Eur päisches Patentamt

Eur p an Pat nt Offic

Office eur pé n des br v ts



EP 1 108 790 A2

(12)

# **EUROPEAN PATENT APPLICATION**

- (43) Date of publication: 20.06.2001 Bulletin 2001/25
- (21) Application number: 00127688.0
- (22) Date of filing: 18.12.2000

- (51) Int CI.7: **C12Q 1/68**, C07H 21/04, C12N 15/63, C07K 14/34, C12R 1/15, G06F 17/00, C12R 1/13, G01N 33/50
- (84) Designated Contracting States:

  AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU

  MC NL PT SE TR

  Designated Extension States:

  AL LT LV MK RO SI
- (30) Priority: 16.12.1999 JP 37748499 07.04.2000 JP 2000159162 03.08.2000 JP 2000280988
- (83) Declaration under Rule 28(4) EPC (expert solution)
- (71) Applicant: KYOWA HAKKO KOGYO CO., LTD. Chiyoda-ku, Tokyo 100-8185 (JP)
- (72) Inventors:
  - Nakagawa, Satochi, c/o Kyowa Hakko Kogyo Co.,Ltd. Machida-shi, Tokyo 194-8533 (JP)
  - Mizoguchi, Hiroshi,
     c/o Kyowa Hakko Kogyo Co.,Ltd.
     Machida-shi, Tokyo 194-8533 (JP)

- Ando, Seiko, c/o Kyowa Hakko Kogyo Co., Ltd. Machida-shi, Tokyo 194-8533 (JP)
- Hayashi, Mikiro, c/o Kyowa Hakko Kogyo Co.,Ltd. Machida-shi, Tokyo 194-8533 (JP)
- Ochiai, Keiko, c/o Kyowa Hakko Kogyo Co.,Ltd. Machida-shi, Tokyo 194-8533 (JP)
- Yokoi, Haruhiko, c/o Kyowa Hakko Kogyo Co.,Ltd. Machida-shi, Tokyo 194-8533 (JP)
- Tateishi, Naoko,
   c/o Kyowa Hakko Kogyo Co.,Ltd.
   Machida-shi, Tokyo 194-8533 (JP)
- Senoh, Akihiro, c/o Kyowa Hakko Kogyo Co.,Ltd. Machida-shi, Tokyo 194-8533 (JP)
- Ikeda, Masato, c/o Kyowa Hakko Kogyo Co.,Ltd. Machida-shi, Tokyo 194-8533 (JP)
- Ozaki, Akio, c/o Kyowa Hakko Kogyo Co., Ltd. Hofu-shi, Yamaguchi 747-8522 (JP)
- (74) Representative: VOSSIUS & PARTNER
  Siebertstrasse 4
  81675 München (DE)

## (54) Novel polynucleotides

(57) Novel polynucleotides derived from microorganisms belonging to coryneform bacteria and fragments thereof, polypeptides encoded by the polynucleotides and fragments thereof, polynucleotide arrays

comprising the polynucleotides and fragments thereof, recording media in which the nucleotide sequences of the polynucleotide and fragments thereof have been recorded which are readable in a computer, and use of them.

**EP 1 108 790 A2** 

### 1. Field of the Invention

[0001] The present invention relates to novel polynucleotides derived from microorganisms belonging to coryneform bacteria and fragments thereof, polypeptides encoded by the polynucleotides and fragments thereof, polypucleotide arrays comprising the polynucleotides and fragments thereof, computer readable recording media in which the nucleotide sequences of the polynucleotide and fragments thereof have been recorded, and use of them as well as a method of using the polynucleotide and/or polypeptide sequence information to make comparisons.

### 2. Brief Description of the Background Art

[0002] Coryneform bacteria are used in producing various useful substances, such as amino acids, nucleic acids, vitamins, saccharides (for example, ribulose), organic acids (for example, pyruvic acid), and analogues of the abovedescribed substances (for example, N-acetylamino acids) and are very useful microorganisms industrially. Many mutants thereof are known.

[0003] For example, Corynebacterium glutamicum is a Gram-positive bacterium identified as a glutamic acid-producing bacterium, and many amino acids are produced by mutants thereof. For example, 1,000,000 ton/year of Lglutamic acid which is useful as a seasoning for umami (delicious taste), 250,000 ton/year of L-lysine which is a valuable additive for livestock feeds and the like, and several hundred ton/year or more of other amino acids, such as L-arginine, L-proline, L-glutamine, L-tryptophan, and the like, have been produced in the world (Nikkei Bio Yearbook 99, published by Nikkei BP (1998)).

[0004] The production of amino acids by Corynebacterium glutamicum is mainly carried out by its mutants (metabolic mutants) which have a mutated metabolic pathway and regulatory systems. In general, an organism is provided with various metabolic regulatory systems so as not to produce more amino acids than it needs. In the biosynthesis of Llysine, for example, a microorganism belonging to the genus Corynebacterium is under such regulation as preventing the excessive production by concerted inhibition by lysine and threonine against the activity of a biosynthesis enzyme common to lysine, threonine and methionine, i.e., an aspartokinase, (J. Biochem., 65: 849-859 (1969)). The biosynthesis of arginine is controlled by repressing the expression of its biosynthesis gene by arginine so as not to biosynthesize an excessive amount of arginine (Microbiology, 142: 99-108 (1996)). It is considered that these metabolic regulatory mechanisms are deregulated in amino acid-producing mutants. Similarly, the metabolic regulation is deregulated in mutants producing nucleic acids, vitamins, saccharides, organic acids and analogues of the above-described substances so as to improve the productivity of the objective product.

[0005] However, accumulation of basic genetic, biochemical and molecular biological data on coryneform bacteria is insufficient in comparison with Escherichia coli, Bacillus subtilis, and the like. Also, few findings have been obtained on mutated genes in amino acid-producing mutants. Thus, there are various mechanisms, which are still unknown, of regulating the growth and metabolism of these microorganisms.

[0006] A chromosomal physical map of Corynebacterium glutamicum ATCC 13032 is reported and it is known that its genome size is about 3,100 kb (Mol. Gen. Genet., 252: 255-265 (1996)). Calculating on the basis of the usual gene density of bacteria, it is presumed that about 3,000 genes are present in this genome of about 3,100 kb. However, only about 100 genes mainly concerning amino acid biosynthesis genes are known in Corynebacterium glutamicum, and the nucleotide sequences of most genes have not been clarified hitherto.

45 [0007] In recent years, the full nucleotide sequence of the genomes of several microorganisms, such as Escherichia coli, Mycobacterium tuberculosis, yeast, and the like, have been determined (Science, 277: 1453-62 (1997); Nature, 393: 537-544 (1998); Nature, 387: 5-105 (1997)). Based on the thus determined full nucleotide sequences, assumption of gene regions and prediction of their function by comparison with the nucleotide sequences of known genes have been carried out. Thus, the functions of a great number of genes have been presumed, without genetic, biochemical 50 or molecular biological experiments.

[0008] In recent years, moreover, techniques for monitoring expression levels of a great number of genes simultaneously or detecting mutations, using DNA chips, DNA arrays or the like in which a partial nucleic acid fragment of a gene or a partial nucleic acid fragment in genomic DNA other than a gene is fixed to a solid support, have been developed. The techniques contribute to the analysis of microorganisms, such as yeasts, Mycobacterium tuberculosis, Mycobacterium bovis used in BCG vaccines, and the like (Science, 278: 680-686 (1997); Proc. Natl. Acad. Sci. USA. 96: 12833-38 (1999); Science, 284: 1520-23 (1999)).

2

10

15

20

25

35

30

55

10

[0009] An object of the present invention is to provide a polynucleotide and a polypeptide derived from a microorganism of coryneform bacteria which are industrially useful, sequence information of the polynucleotide and the polypeptide, a method for analyzing the microorganism, an apparatus and a system for use in the analysis, and a method for breeding the microorganism.

[0010] The present invention provides a polynucleotide and an oligonucleotide derived from a microorganism belonging to coryneform bacteria, oligonucleotide arrays to which the polynucleotides and the oligonucleotides are fixed, a polypeptide encoded by the polynucleotide, an antibody which recognizes the polypeptide, polypeptide arrays to which the polypeptides or the antibodies are fixed, a computer readable recording medium in which the nucleotide sequences of the polynucleotide and the oligonucleotide and the amino acid sequence of the polypeptide have been recorded, and a system based on the computer using the recording medium as well as a method of using the polynucleotide and/or polypeptide sequence information to make comparisons.

15 BRIEF DESCRIPTION OF THE DRAWING

> [0011] Fig. 1 is a map showing the positions of typical genes on the genome of Corynebacterium glutamicum ATCC 13032.

> [0012] Fig. 2 is electrophoresis showing the results of proteome analyses using proteins derived from (A) Corynebacterium glutamicum ATCC 13032, (B) FERM BP-7134, and (C) FERM BP-158.

> [0013] Fig. 3 is a flow chart of an example of a system using the computer readable media according to the present invention.

> [0014] Fig. 4 is a flow chart of an example of a system using the computer readable media according to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0015] This application is based on Japanese applications No. Hei. 11-377484 filed on December 16, 1999, No. 2000-159162 filed on April 7, 2000 and No. 2000-280988 filed on August 3, 2000, the entire contents of which are incorporated hereinto by reference.

[0016] From the viewpoint that the determination of the full nucleotide sequence of Corynebacterium glutamicum would make it possible to specify gene regions which had not been previously identified, to determine the function of an unknown gene derived from the microorganism through comparison with nucleotide sequences of known genes and amino acid sequences of known genes, and to obtain a useful mutant based on the presumption of the metabolic regulatory mechanism of a useful product by the microorganism, the inventors conducted intensive studies and, as a result, found that the complete genome sequence of Corynebacterium glutamicum can be determined by applying the whole genome shotgun method.

[0017] Specifically, the present invention relates to the following (1) to (65):

- (1) A method for at least one of the following:
  - (A) identifying a mutation point of a gene derived from a mutant of a coryneform bacterium,
  - (B) measuring an expression amount of a gene derived from a coryneform bacterium,
  - (C) analyzing an expression profile of a gene derived from a coryneform bacterium,
  - (D) analyzing expression patterns of genes derived from a coryneform bacterium, or
  - (E) identifying a gene homologous to a gene derived from a coryneform bacterium, said method comprising:
    - (a) producing a polynucleotide array by adhering to a solid support at least two polynucleotides selected from the group consisting of first polynucleotides comprising the nucleotide sequence represented by any one of SEQ ID NOS:1 to 3501, second polynucleotides which hybridize with the first polynucleotides under stringent conditions, and third polynucleotides comprising a sequence of 10 to 200 continuous bases of the first or second polynucleotides,
    - (b) incubating the polynucleotide array with at least one of a labeled polynucleotide derived from a coryneform bacterium, a labeled polynucleotide derived from a mutant of the coryneform bacterium or a labeled polynucleotide to be examined, under hybridization conditions,
    - (c) detecting any hybridization, and
    - (d) analyzing the result of the hybridization.

20

25

30

35

40

45

50

55

As used herein, for example, the at least two polynucleotides can be at least two of the first polynucleotides, at least two of the second polynucleotides, at least two of the third polynucleotides, or at least two of the first, second and third polynucleotides.

- (2) The method according to (1), wherein the coryneform bacterium is a microorganism belonging to the genus Corynebacterium, the genus Brevibacterium, or the genus Microbacterium.
  - (3) The method according to (2), wherein the microorganism belonging to the genus Corynebacterium is selected from the group consisting of Corynebacterium glutamicum, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum, Corynebacterium callunae, Corynebacterium herculis, Corynebacterium lilium, Corynebacterium melassecola, Corynebacterium thermoaminogenes, and Corynebacterium ammoniagenes.
  - (4) The method according to (1), wherein the polynucleotide derived from a coryneform bacterium, the polynucleotide derived from a mutant of the coryneform bacterium or the polynucleotide to be examined is a gene relating to the biosynthesis of at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogues thereof.
  - (5) The method according to (1), wherein the polynucleotide to be examined is derived from Escherichia coli.
  - (6) A polynucleotide array, comprising:

5

10

15

20

25

30

35

40

45

50

55

at least two polynucleotides selected from the group consisting of first polynucleotides comprising the nucleotide sequence represented by any one of SEQ ID NOS:1 to 3501, second polynucleotides which hybridize with the first polynucleotides under stringent conditions, and third polynucleotides comprising 10 to 200 continuous bases of the first or second polynucleotides, and a solid support adhered thereto.

As used herein, for example, the at least two polynucleotides can be at least two of the first polynucleotides, at least two of the second polynucleotides, at least two of the third polynucleotides, or at least two of the first, second and third polynucleotides.

- (7) A polynucleotide comprising the nucleotide sequence represented by SEQ ID NO:1 or a polynucleotide having a homology of at least 80% with the polynucleotide.
- (8) A polynucleotide comprising any one of the nucleotide sequences represented by SEQ ID NOS:2 to 3431, or a polynucleotide which hybridizes with the polynucleotide under stringent conditions.
- (9) A polynucleotide encoding a polypeptide having any one of the amino acid sequences represented by SEQ ID NOS:3502 to 6931, or a polynucleotide which hybridizes therewith under stringent conditions.
- (10) A polynucleotide which is present in the 5' upstream or 3' downstream of a polynucleotide comprising the nucleotide sequence of any one of SEQ ID NOS:2 to 3431 in a whole polynucleotide comprising the nucleotide sequence represented by SEQ ID NO:1, and has an activity of regulating an expression of the polynucleotide.
- (11) A polynucleotide comprising 10 to 200 continuous bases in the nucleotide sequence of the polynucleotide of any one of (7) to (10), or a polynucleotide comprising a nucleotide sequence complementary to the polynucleotide comprising 10 to 200 continuous based.
- (12) A recombinant DNA comprising the polynucleotide of any one of (8) to (11).
- (13) A transformant comprising the polynucleotide of any one of (8) to (11) or the recombinant DNA of (12).
- (14) A method for producing a polypeptide, comprising:

culturing the transformant of (13) in a medium to produce and accumulate a polypeptide encoded by the polynucleotide of (8) or (9) in the medium, and recovering the polypeptide from the medium.

- (15) A method for producing at least one of an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogues thereof, comprising:
  - culturing the transformant of (13) in a medium to produce and accumulate at least one of an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogues thereof in the medium, and recovering the at least one of the amino acid, the nucleic acid, the vitamin, the saccharide, the organic acid, and analogues thereof from the medium.
- (16) A polypeptide encoded by a polynucleotide comprising the nucleotide sequence selected from SEQ ID NOS: 2 to 3431.
- (17) A polypeptide comprising the amino acid sequence selected from SEQ ID NOS:3502 to 6931.
- (18) The polypeptide according to (16) or (17), wherein at least one amino acid is deleted, replaced, inserted or

added, said polypeptides having an activity which is substantially the same as that of the polypeptide without said at least one amino acid deletion, replacement, insertion or addition.

- (19) A polypeptide comprising an amino acid sequence having a homology of at least 60% with the amino acid sequence of the polypeptide of (16) or (17), and having an activity which is substantially the same as that of the polypeptide.
- (20) An antibody which recognizes the polypeptide of any one of (16) to (19).
- (21) A polypeptide array, comprising:

5

10

15

20

25

30

35

40

45

50

55

at least one polypeptide or partial fragment polypeptide selected from the polypeptides of (16) to (19) and partial fragment polypeptides of the polypeptides, and a solid support adhered thereto.

(22) A polypeptide array, comprising:

at least one antibody which recognizes a polypeptide or partial fragment polypeptide selected from the polypeptides of (16) to (19) and partial fragment polypeptides of the polypeptides, and a solid support adhered thereto.

- (23) A system based on a computer for identifying a target sequence or a target structure motif derived from a coryneform bacterium, comprising the following:
  - (i) a user input device that inputs at least one nucleotide sequence information selected from SEQ ID NOS:1 to 3501, and target sequence or target structure motif information;
  - (ii) a data storage device for at least temporarily storing the input information;
  - (iii) a comparator that compares the at least one nucleotide sequence information selected from SEQ ID NOS: 1 to 3501 with the target sequence or target structure motif information, recorded by the data storage device for screening and analyzing nucleotide sequence information which is coincident with or analogous to the target sequence or target structure motif information; and
  - (iv) an output device that shows a screening or analyzing result obtained by the comparator.
- (24) A method based on a computer for identifying a target sequence or a target structure motif derived from a coryneform bacterium, comprising the following:
  - (i) inputting at least one nucleotide sequence information selected from SEQ ID NOS:1 to 3501, target sequence information or target structure motif information into a user input device;
  - (ii) at least temporarily storing said information;
  - (iii) comparing the at least one nucleotide sequence information selected from SEQ ID NOS:1 to 3501 with the target sequence or target structure motif information; and
  - (iv) screening and analyzing nucleotide sequence information which is coincident with or analogous to the target sequence or target structure motif information.
- (25) A system based on a computer for identifying a target sequence or a target structure motif derived from a coryneform bacterium, comprising the following:
  - (i) a user input device that inputs at least one amino acid sequence information selected from SEQ ID NOS: 3502 to 7001, and target sequence or target structure motif information;
  - (ii) a data storage device for at least temporarily storing the input information;
  - (iii) a comparator that compares the at least one amino acid sequence information selected from SEQ ID NOS: 3502 to 7001 with the target sequence or target structure motif information, recorded by the data storage device for screening and analyzing amino acid sequence information which is coincident with or analogous to the target sequence or target structure motif information; and
  - (iv) an output device that shows a screening or analyzing result obtained by the comparator.
- (26) A method based on a computer for identifying a target sequence or a target structure motif derived from a coryneform bacterium, comprising the following:
  - (i) inputting at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001, and target sequence information or target structure motif information into a user input device;

(ii) at least temporarily storing said information;

5

10

15

25

30

35

40

45

50

55

- (iii) comparing the at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001 with the target sequence or target structure motif information; and
- (iv) screening and analyzing amino acid sequence information which is coincident with or analogous to the target sequence or target structure motif information.
- (27) A system based on a computer for determining a function of a polypeptide encoded by a polynucleotide having a target nucleotide sequence derived from a coryneform bacterium, comprising the following:
  - (i) a user input device that inputs at least one nucleotide sequence information selected from SEQ ID NOS:2 to 3501, function information of a polypeptide encoded by the nucleotide sequence, and target nucleotide sequence information;
  - (ii) a data storage device for at least temporarily storing the input information;
  - (iii) a comparator that compares the at least one nucleotide sequence information selected from SEQ ID NOS: 2 to 3501 with the target nucleotide sequence information, and determining a function of a polypeptide encoded by a polynucleotide having the target nucleotide sequence which is coincident with or analogous to the polynucleotide having at least one nucleotide sequence selected from SEQ ID NOS:2 to 3501; and
  - (iv) an output devices that shows a function obtained by the comparator.
- (28) A method based on a computer for determining a function of a polypeptide encoded by a polypeptide encoded by a polynucleotide having a target nucleotide sequence derived from a coryneform bacterium, comprising the following:
  - (i) inputting at least one nucleotide sequence information selected from SEQ ID NOS:2 to 3501, function information of a polypeptide encoded by the nucleotide sequence, and target nucleotide sequence information; (ii) at least temporarily storing said information;
  - (iii) comparing the at least one nucleotide sequence information selected from SEQ ID NOS:2 to 3501 with the target nucleotide sequence information; and
  - (iv) determining a function of a polypeptide encoded by a polynucleotide having the target nucleotide sequence which is coincident with or analogous to the polynucleotide having at least one nucleotide sequence selected from SEQ ID NOS:2 to 3501.
  - (29) A system based on a computer for determining a function of a polypeptide having a target amino acid sequence derived from a coryneform bacterium, comprising the following:
    - (i) a user input device that inputs at least one amino acid sequence information selected from SEQ ID NOS: 3502 to 7001, function information based on the amino acid sequence, and target amino acid sequence information:
    - (ii) a data storing device for at least temporarily storing the input information;
    - (iii) a comparator that compares the at least one amino acid sequence information selected from SEQ ID NOS: 3502 to 7001 with the target amino acid sequence information for determining a function of a polypeptide having the target amino acid sequence which is coincident with or analogous to the polypeptide having at least one amino acid sequence selected from SEQ ID NOS:3502 to 7001; and
    - (iv) an output device that shows a function obtained by the comparator.
  - (30) A method based on a computer for determining a function of a polypeptide having a target amino acid sequence derived from a coryneform bacterium, comprising the following:
    - (i) inputting at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001, function information based on the amino acid sequence, and target amino acid sequence information;
    - (ii) at least temporarily storing said information;
    - (iii) comparing the at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001 with the target amino acid sequence information; and
    - (iv) determining a function of a polypeptide having the target amino acid sequence which is coincident with or analogous to the polypeptide having at least one amino acid sequence selected from SEQ ID NOS:3502 to 7001.
  - (31) The system according to any one of (23), (25), (27) and (29), wherein a coryneform bacterium is a microor-

ganism of the genus Corynebacterium, the genus Brevibacterium, or the genus Microbacterium.

- (32) The method according to any one of (24), (26), (28) and (30), wherein a coryneform bacterium is a microorganism of the genus *Corynebacterium*, the genus *Brevibacterium*, or the genus *Microbacterium*.
- (33) The system according to (31), wherein the microorganism belonging to the genus *Corynebacterium* is selected from the group consisting of *Corynebacterium glutamicum*, *Corynebacterium acetoacidophilum*, *Corynebacterium acetoglutamicum*, *corynebacterium callunae*, *corynebacterium herculis*, *Corynebacterium lilium*, *Corynebacterium melassecola*, *Corynebacterium thermoaminogenes*, and *Corynebacterium ammoniagenes*.
- (34) The method according to (32), wherein the microorganism belonging to the genus *Corynebacterium* is selected from the group consisting of *Corynebacterium glutamicum*, *Corynebacterium acetoacidophilum*, *Corynebacterium acetoglutamicum*, *Corynebacterium callunae*, *Corynebacterium herculis*, *Corynebacterium lilium*, *Corynebacterium melassecola*, *Corynebacterium thermoaminogenes*, and *Corynebacterium ammoniagenes*.
- (35) A recording medium or storage device which is readable by a computer in which at least one nucleotide sequence information selected from SEQ ID NOS:1 to 3501 or function information based on the nucleotide sequence is recorded, and is usable in the system of (23) or (27) or the method of (24) or (28).
- (36) A recording medium or storage device which is readable by a computer in which at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001 or function information based on the amino acid sequence is recorded, and is usable in the system of (25) or (29) or the method of (26) or (30).
- (37) The recording medium or storage device according to

5

10

15

20

25

30

35

40

45

50

55

- (35) or (36), which is a computer readable recording medium selected from the group consisting of a floppy disc, a hard disc, a magnetic tape, a random access memory (RAM), a read only memory (ROM), a magneto-optic disc (MO), CD-ROM, CD-R, CD-RW, DVD-ROM, DVD-RAM and DVD-RW.
- (38) A polypeptide having a homoserine dehydrogenase activity, comprising an amino acid sequence in which the Val residue at the 59th in the amino acid sequence of homoserine dehydrogenase derived from a coryneform bacterium is replaced with an amino acid residue other than a Val residue.
- (39) A polypeptide comprising an amino acid sequence in which the Val residue at the 59th position in the amino acid sequence as represented by SEQ ID NO:6952 is replaced with an amino acid residue other than a Val residue.
- (40) The polypeptide according to (38) or (39), wherein the Val residue at the 59th position is replaced with an Ala residue.
- (41) A polypeptide having pyruvate carboxylase activity, comprising an amino acid sequence in which the Pro residue at the 458th position in the amino acid sequence of pyruvate carboxylase derived from a coryneform bacterium is replaced with an amino acid residue other than a Pro residue.
- (42) A polypeptide comprising an amino acid sequence in which the Pro residue at the 458th position in the amino acid sequence represented by SEQ ID NO:4265 is replaced with an amino acid residue other than a Pro residue.
- (43) The polypeptide according to (41) or (42), wherein the Pro residue at the 458th position is replaced with a Ser residue.
- (44) The polypeptide according to any one of (38) to (43), which is derived from Corynebacterium glutamicum.
- (45) A DNA encoding the polypeptide of any one of (38) to (44).
- (46) A recombinant DNA comprising the DNA of (45).
- (47) A transformant comprising the recombinant DNA of (46).
- (48) A transformant comprising in its chromosome the DNA of (45).
- (49) The transformant according to (47) or (48), which is derived from a coryneform bacterium.
- (50) The transformant according to (49), which is derived from Corynebacterium glutamicum.
- (51) A method for producing L-lysine, comprising:

culturing the transformant of any one of (47) to (50) in a medium to produce and accumulate L-lysine in the medium, and

recovering the L-lysine from the culture.

- (52) A method for breeding a coryneform bacterium using the nucleotide sequence information represented by SEQ ID NOS:1 to 3431, comprising the following:
  - (i) comparing a nucleotide sequence of a genome or gene of a production strain derived a coryneform bacterium which has been subjected to mutation breeding so as to produce at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogous thereof by a fermentation method, with a corresponding nucleotide sequence in SEQ ID NOS:1 to 3431;
  - (ii) identifying a mutation point present in the production strain based on a result obtained by (i);
  - (iii) introducing the mutation point into a coryneform bacterium which is free of the mutation point; and
  - (iv) examining productivity by the fermentation method of the compound selected in (i) of the coryneform

bacterium obtained in (iii).

5

10

15

20

25

30

35

40

45

50

55

- (53) The method according to (52), wherein the gene is a gene encoding an enzyme in a biosynthetic pathway or a signal transmission pathway.
- (54) The method according to (52), wherein the mutation point is a mutation point relating to a useful mutation which improves or stabilizes the productivity.
- (55) A method for breading a coryneform bacterium using the nucleotide sequence information represented by SEQ ID NOS:1 to 3431, comprising:
  - (i) comparing a nucleotide sequence of a genome or gene of a production strain derived a coryneform bacterium which has been subjected to mutation breeding so as to produce at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogous thereof by a fermentation method, with a corresponding nucleotide sequence in SEQ ID NOS:1 to 3431;
  - (ii) identifying a mutation point present in the production strain based on a result obtain by (i);
  - (iii) deleting a mutation point from a coryneform bacterium having the mutation point; and
  - (iv) examining productivity by the fermentation method of the compound selected in (i) of the coryneform bacterium obtained in (iii).
- (56) The method according to (55), wherein the gene is a gene encoding an enzyme in a biosynthetic pathway or a signal transmission pathway.
- (57) The method according to (55), wherein the mutation point is a mutation point which decreases or destabilizes the productivity.
- (58) A method for breeding a coryneform bacterium using the nucleotide sequence information represented by SEQ ID NOS:2 to 3431, comprising the following:
  - (i) identifying an isozyme relating to biosynthesis of at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogous thereof, based on the nucleotide sequence information represented by SEQ ID NOS:2 to 3431;
  - (ii) classifying the isozyme identified in (i) into an isozyme having the same activity;
  - (iii) mutating all genes encoding the isozyme having the same activity simultaneously; and
  - (iv) examining productivity by a fermentation method of the compound selected in (i) of the coryneform bacterium which have been transformed with the gene obtained in (iii).
- (59) A method for breeding a coryneform bacterium using the nucleotide sequence information represented by SEQ ID NOS:2 to 3431, comprising the following:
  - (i) arranging a function information of an open reading frame (ORF) represented by SEQ ID NOS:2 to 3431;
  - (ii) allowing the arranged ORF to correspond to an enzyme on a known biosynthesis or signal transmission pathway;
  - (iii) explicating an unknown biosynthesis pathway or signal transmission pathway of a coryneform bacterium in combination with information relating known biosynthesis pathway or signal transmission pathway of a coryneform bacterium;
  - (iv) comparing the pathway explicated in (iii) with a biosynthesis pathway of a target useful product; and
  - (v) transgenetically varying a coryneform bacterium based on the nucleotide sequence information to either strengthen a pathway which is judged to be important in the biosynthesis of the target useful product in (iv) or weaken a pathway which is judged not to be important in the biosynthesis of the target useful product in (iv).
- (60) A coryneform bacterium, bred by the method of any one of (52) to (59).
- (61) The coryneform bacterium according to (60), which is a microorganism belonging to the genus *Corynebacterium*, the genus *Brevibacterium*, or the genus *Microbacterium*.
- (62) The coryneform bacterium according to (61), wherein the microorganism belonging to the genus Corynebacterium is selected from the group consisting of Corynebacterium glutamicum, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum, Corynebacterium callunae, Corynebacterium herculis, Corynebacterium lilium, Corynebacterium melassecola, Corynebacterium thermoaminogenes, and Corynebacterium ammoniagenes.
- (63) A method for producing at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid and an analogue thereof, comprising:

culturing a coryneform bacterium of any one of (60) to (62) in a medium to produce and accumulate at least

one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogues thereof;

recovering the compound from the culture.

- (64) The method according to (63), wherein the compound is L-lysine.
- (65) A method for identifying a protein relating to useful mutation based on proteome analysis, comprising the following:
  - (i) preparing

5

10

15

20

25

30

45

50

55

a protein derived from a bacterium of a production strain of a coryneform bacterium which has been subjected to mutation breeding by a fermentation process so as to produce at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogues thereof, and a protein derived from a bacterium of a parent strain of the production strain;

- (ii) separating the proteins prepared in (i) by two dimensional electrophoresis;
- (iii) detecting the separated proteins, and comparing an expression amount of the protein derived from the production strain with that derived from the parent strain;
- (iv) treating the protein showing different expression amounts as a result of the comparison with a peptidase to extract peptide fragments;
- (v) analyzing amino acid sequences of the peptide fragments obtained in (iv); and
- (vi) comparing the amino acid sequences obtained in (v) with the amino acid sequence represented by SEQ ID NOS:3502 to 7001 to identifying the protein having the amino acid sequences.

As used herein, the term "proteome", which is a coined word by combining "protein" with "genome", refers to a method for examining of a gene at the polypeptide level.

- (66) The method according to (65), wherein the coryneform bacterium is a microorganism belonging to the genus *Corynebacterium*, the genus *Brevibacterium*, or the genus *Microbacterium*.
- (67) The method according to (66), wherein the microorganism belonging to the genus Corynebacterium is selected from the group consisting of Corynebacterium glutamicum, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum, Corynebacterium callunae, corynebacterium herculis, Corynebacterium lilium Corynebacterium melassecola, Corynebacterium thermoaminogenes, and Corynebacterium ammoniagenes.
- (68) A biologically pure culture of Corynebacterium glutamicum AHP-3 (FERM BP-7382).
- 35 [0018] The present invention will be described below in more detail, based on the determination of the full nucleotide sequence of coryneform bacteria.
  - 1. Determination of full nucleotide sequence of coryneform bacteria
- [0019] The term "coryneform bacteria" as used herein means a microorganism belonging to the genus Corynebacterium, the genus Brevibacterium or the genus Microbacterium as defined in Bergeys Manual of Determinative Bacteriology, 8: 599 (1974).
  - [0020] Examples include Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum, Corynebacterium callunae, Corynebacterium glutamicum, Corynebacterium herculis, Corynebacterium lilium, Corynebacterium melassecola, Corynebacterium thermoaminogenes, Brevibacterium saccharolyticum, Brevibacterium immariophilum, Brevibacterium roseum, Brevibacterium thiogenitalis, Microbacterium ammoniaphilum, and the like.
  - [0021] Specific examples include Corynebacterium acetoacidophilum ATCC 13870, Corynebacterium acetoglutamicum ATCC 15806, Corynebacterium callunae ATCC 15991, Corynebacterium glutamicum ATCC 13032, Corynebacterium glutamicum ATCC 13060, Corynebacterium glutamicum ATCC 13826 (prior genus and species: Brevibacterium flavum, or Corynebacterium lactofermentum), Corynebacterium glutamicum ATCC 14020 (prior genus and species: Brevibacterium divaricatum), Corynebacterium glutamicum ATCC 13869 (prior genus and species: Brevibacterium lactofermentum), Corynebacterium herculis ATCC 13868, Corynebacterium lilium ATCC 15990, Corynebacterium melassecola ATCC 17965, Corynebacterium thermoaminogenes FERM 9244, Brevibacterium saccharolyticum ATCC 14066, Brevibacterium immariophilum ATCC 14068, Brevibacterium roseum ATCC 13825, Brevibacterium thiogenitalis ATCC 19240, Microbacterium ammoniaphilum ATCC 15354, and the like.

## (1) Preparation of genome DNA of coryneform bacteria

5

10

15

20

25

30

35

40

45

50

55

[0022] Coryneform bacteria can be cultured by a conventional method.

[0023] Any of a natural medium and a synthetic medium can be used, so long as it is a medium suitable for efficient culturing of the microorganism, and it contains a carbon source, a nitrogen source, an inorganic salt, and the like which can be assimilated by the microorganism.

[0024] In Corynebacterium glutamicum, for example, a BY medium (7 g/l meat extract, 10 g/l peptone, 3 g/l sodium chloride, 5 g/l yeast extract, pH 7.2) containing 1% of glycine and the like can be used. The culturing is carried out at 25 to 35°C overnight.

[0025] After the completion of the culture, the cells are recovered from the culture by centrifugation. The resulting cells are washed with a washing solution.

[0026] Examples of the washing solution include STE buffer (10.3% sucrose, 25 mmol/l Tris hydrochloride, 25 mmol/l ethylenediaminetetraacetic acid (hereinafter referred to as "EDTA"), pH 8.0), and the like.

[0027] Genome DNA can be obtained from the washed cells according to a conventional method for obtaining genome DNA, namely, lysing the cell wall of the cells using a lysozyme and a surfactant (SDS, etc.), eliminating proteins and the like using a phenol solution and a phenol/chloroform solution, and then precipitating the genome DNA with ethanol or the like. Specifically, the following method can be illustrated.

[0028] The washed cells are suspended in a washing solution containing 5 to 20 mg/l lysozyme. After shaking, 5 to 20% SDS is added to lyse the cells. In usual, shaking is gently performed at 25 to 40°C for 30 minutes to 2 hours. After shaking, the suspension is maintained at 60 to 70°C for 5 to 15 minutes for the lysis.

[0029] After the lysis, the suspension is cooled to ordinary temperature, and 5 to 20 ml of Tris-neutralized phenol is added thereto, followed by gently shaking at room temperature for 15 to 45 minutes.

[0030] After shaking, centrifugation (15,000  $\times$  g, 20 minutes, 20°C) is carried out to fractionate the aqueous layer.

[0031] After performing extraction with phenol/chloroform and extraction with chloroform (twice) in the same manner, 3 mol/l sodium acetate solution (pH 5.2) and isopropanol are added to the aqueous layer at 1/10 times volume and 2 times volume, of the aqueous layer, respectively, followed by gently stirring to precipitate the genome DNA.

[0032] The genome DNA is dissolved again in a buffer containing 0.01 to 0.04 mg/ml RNase. As an example of the buffer, TE buffer (10 mmol/l Tris hydrochloride, 1 mol/l EDTA, pH 8.0) can be used. After dissolving, the resultant solution is maintained at 25 to 40°C for 20 to 50 minutes and then extracted successively with phenol, phenol/chloroform and chloroform as in the above case.

[0033] After the extraction, isopropanol precipitation is carried out and the resulting DNA precipitate is washed with 70% ethanol, followed by air drying, and then dissolved in TE buffer to obtain a genome DNA solution.

## (2) Production of shotgun library

[0034] A method for produce a genome DNA library using the genome DNA of the coryneform bacteria prepared in the above (1) include a method described in *Molecular Cloning, A laboratory Manual,* Second Edition (1989) (hereinafter referred to as "*Molecular Cloning,* 2nd ed."). In particular, the following method can be exemplified to prepare a genome DNA library appropriately usable in determining the full nucleotide sequence by the shotgun method.

[0035] To 0.01 mg of the genome DNA of the coryneform bacteria prepared in the above (1), a buffer, such as TE buffer or the like, is added to give a total volume of 0.4 ml. Then, the genome DNA is digested into fragments of 1 to 10 kb with a sonicator (Yamato Powersonic Model 50). The treatment with the sonicator is performed at an output of 20 continuously for 5 seconds.

[0036] The resulting genome DNA fragments are blunt-ended using DNA blunting kit (manufactured by Takara Shuzo) or the like.

[0037] The blunt-ended genome fragments are fractionated by agarose gel or polyacrylamide gel electrophoresis and genome fragments of 1 to 2 kb are cut out from the gel.

[0038] To the gel, 0.2 to 0.5 ml of a buffer for eluting DNA, such as MG elution buffer (0.5 mol/l ammonium acetate, 10 mmol/l magnesium acetate, 1 mmol/l EDTA, 0.1% SDS) or the like, is added, followed by shaking at 25 to 40°C overnight to elute DNA.

[0039] The resulting DNA eluate is treated with phenol/chloroform and then precipitated with ethanol to obtain a genome library insert.

[0040] This insert is ligated into a suitable vector, such as pUC18 Smal/SAP (manufactured by Amersham Pharmacia Biotech) or the like, using T4 ligase (manufactured by Takara Shuzo) or the like. The ligation can be carried out by allowing a mixture to stand at 10 to 20°C for 20 to 50 hours.

[0041] The resulting ligation product is precipitated with ethanol and dissolved in 5 to 20 μl of TE buffer.

[0042] Escherichia coli is transformed in accordance with a conventional method using 0.5 to 2  $\mu$ I of the ligation solution. Examples of the transformation method include the electroporation method using ELECTRO MAX DHIOB

(manufactured by Life Technologies) for *Escherichia coli*. The electroporation method can be carried out under the conditions as described in the manufacturer's instructions.

[0043] The transformed *Escherichia coli* is spread on a suitable selection medium containing agar, for example, LB plate medium containing 10 to 100 mg/l ampicillin (LB medium (10 g/l bactotrypton, 5 g/l yeast extract, 10 g/l sodium chloride, pH 7.0) containing 1.6% of agar) when pUC18 is used as the cloning vector, and cultured therein.

[0044] The transformant can be obtained as colonies formed on the plate medium. In this step, it is possible to select the transformant having the recombinant DNA containing the genome DNA as white colonies by adding X-gal and IPTG (isopropyl-β-thiogalactopyranoside) to the plate medium.

[0045] The transformant is allowed to stand for culturing in a 96-well titer plate to which 0.05 ml of the LB medium containing 0.1 mg/ml of ampicillin has been added in each well. The resulting culture can be used in an experiment of (4) described below. Also, the culture solution can be stored at -80°C by adding 0.05 ml per well of the LB medium containing 20% glycerol to the culture solution, followed by mixing, and the stored culture solution can be used at any time.

### (3) Production of cosmid library

5

10

15

20

25

30

35

40

45

50

55

[0046] The genome DNA (0.1 mg) of the coryneform bacteria prepared in the above (1) is partially digested with a restriction enzyme, such as *Sau*3Al or the like, and then ultracentrifuged (26,000 rpm, 18 hours, 20°C) under a 10 to 40% sucrose density gradient using a 10% sucrose buffer (1 mol/l Nacl, 20 mmol/l Tris hydrochloride, 5 mmol/l EDTA, 10% sucrose, pH 8.0) and a 40% sucrose buffer (elevating the concentration of the 10% sucrose buffer to 40%).

[0047] After the centrifugation, the thus separated solution is fractionated into tubes in 1 ml per each tube. After confirming the DNA fragment size of each fraction by agarose gel electrophoresis, a fraction rich in DNA fragments of about 40 kb is precipitated with ethanol.

[0048] The resulting DNA fragment is ligated to a cosmid vector having a cohesive end which can be ligated to the fragment. When the genome DNA is partially digested with Sau3AI, the partially digested product can be ligated to, for example, the BamHI site of superCos1 (manufactured by Stratagene) in accordance with the manufacture's instructions.

[0049] The resulting ligation product is packaged using a packaging extract which can be prepared by a method described in *Molecular Cloning*, 2nd ed. and then used in transforming *Escherichia coli*. More specifically, the ligation product is packaged using, for example, a commercially available packaging extract, Gigapack III Gold Packaging Extract (manufactured by Stratagene) in accordance with the manufacture's instructions and then introduced into *Escherichia coli* XL-1-BlueMR (manufactured by Stratagene) or the like.

[0050] The thus transformed *Escherichia coli is* spread on an LB plate medium containing ampicillin, and cultured therein.

[0051] The transformant can be obtained as colonies formed on the plate medium.

[0052] The transformant is subjected to standing culture in a 96-well titer plate to which 0.05 ml of the LB medium containing 0.1 mg/ml ampicillin has been added.

[0053] The resulting culture can be employed in an experiment of (4) described below. Also, the culture solution can be stored at -80°C by adding 0.05 ml per well of the LB medium containing 20% glycerol to the culture solution, followed by mixing, and the stored culture solution can be used at any time.

### (4) Determination of nucleotide sequence

### (4-1) Preparation of template

[0054] The full nucleotide sequence of genome DNA of coryneform bacteria can be determined basically according to the whole genome shotgun method (Science, 269: 496-512 (1995)).

[0055] The template used in the whole genome shotgun method can be prepared by PCR using the library prepared in the above (2) (DNA Research, 5: 1-9 (1998)).

[0056] Specifically, the template can be prepared as follows.

[0057] The clone derived from the whole genome shotgun library is inoculated by using a replicator (manufactured by GENETIX) into each well of a 96-well plate to which 0.08 ml per well of the LB medium containing 0.1 mg/ml ampicillin has been added, followed by stationarily culturing at 37°C overnight.

[0058] Next, the culture solution is transported, using a copy plate (manufactured by Tokken), into each well of a 96-well reaction plate (manufactured by PE Biosystems) to which 0.025 ml per well of a PCR reaction solution has been added using TaKaRa Ex Taq (manufactured by Takara Shuzo). Then, PCR is carried out in accordance with the protocol by Makino et al. (DNA Research, 5: 1-9 (1998)) using GeneAmp PCR System 9700 (manufactured by PE Biosystems) to amplify the inserted fragments.

[0059] The excessive primers and nucleotides are eliminated using a kit for purifying a PCR product, and the product is used as the template in the sequencing reaction.

[0060] It is also possible to determine the nucleotide sequence using a double-stranded DNA plasmid as a template.

[0061] The double-stranded DNA plasmid used as the template can be obtained by the following method.

[0062] The clone derived from the whole genome shotgun library is inoculated into each well of a 24- or 96-well plate to which 1.5 ml per well of a 2 × YT medium (16 g/l bactotrypton, 10 g/l yeast extract, 5 g/l sodium chloride, pH 7.0) containing 0.05 mg/ml ampicillin has been added, followed by culturing under shaking at 37°C overnight.

[0063] The double-stranded DNA plasmid can be prepared from the culture solution using an automatic plasmid preparing machine KURABO PI-50 (manufactured by Kurabo Industries), a multiscreen (manufactured by Millipore) or the like, according to each protocol.

[0064] To purify the plasmid, Biomek 2000 manufactured by Beckman Coulter and the like can be used.

[0065] The resulting purified double-stranded DNA plasmid is dissolved in water to give a concentration of about 0.1 mg/ml. Then, it can be used as the template in sequencing.

### 15 (4-2) Sequencing reaction

5

10

20

25

35

40

45

50

55

[0066] The sequencing reaction can be carried out according to a commercially available sequence kit or the like. A specific method is exemplified below.

[0067] To 6 µl of a solution of ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (manufactured by PE Biosystems), 1 to 2 pmol of an M13 regular direction primer (M13-21) or an M13 reverse direction primer (MI3REV) (DNA Research, 5: 1-9 (1998)) and 50 to 200 ng of the template prepared in the above (4-1) (the PCR product or plasmid) to give 10 µl of a sequencing reaction solution.

[0068] A dye terminator sequencing reaction (35 to 55 cycles) is carried out using this reaction solution and GeneAmp PCR System 9700 (manufactured by PE Biosystems) or the like. The cycle parameter can be determined in accordance with a commercially available kit, for example, the manufacture's instructions attached with ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction Kit.

[0069] The sample can be purified using a commercially available product, such as Multi Screen HV plate (manufactured by Millipore) or the like, according to the manufacture's instructions.

[0070] The thus purified reaction product is precipitated with ethanol, dried and then used for the analysis. The dried reaction product can be stored in the dark at -30°C and the stored reaction product can be used at any time.

[0071] The dried reaction product can be analyzed using a commercially available sequencer and an analyzer according to the manufacture's instructions.

[0072] Examples of the commercially available sequencer include ABI PRISM 377 DNA Sequencer (manufactured by PE Biosystems). Example of the analyzer include ABI PRISM 3700 DNA Analyzer (manufactured by PE Biosystems).

## (5) Assembly

[0073] A software, such as phred (The University of Washington) or the like, can be used as base call for use in analyzing the sequence information obtained in the above (4). A software, such as Cross\_Match (The University of Washington) or SPS Cross\_Match (manufactured by Southwest Parallel Software) or the like, can be used to mask the vector sequence information.

[0074] For the assembly, a software, such as phrap (The University of Washington), SPS phrap (manufactured by Southwest Parallel Software) or the like, can be used.

[0075] In the above, analysis and output of the results thereof, a computer such as UNIX, PC, Macintosh, and the like can be used.

[0076] Contig obtained by the assembly can be analyzed using a graphical editor such as consed (The University of Washington) or the like.

[0077] It is also possible to perform a series of the operations from the base call to the assembly in a lump using a script phredPhrap attached to the consed.

[0078] As used herein, software will be understood to also be referred to as a comparator.

## (6) Determination of nucleotide sequence in gap part

[0079] Each of the cosmids in the cosmid library constructed in the above (3) is prepared in the same manner as in the preparation of the double-stranded DNA plasmid described in the above (4-1). The nucleotide sequence at the end of the insert fragment of the cosmid is determined using a commercially available kit, such as ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (manufactured by PE Biosystems) according to the manufacture's instructions.

[0080] About 800 cosmid clones are sequenced at both ends of the inserted fragment to detect a nucleotide sequence in the contig derived from the shotgun sequencing obtained in (5) which is coincident with the sequence. Thus, the chain linkage between respective cosmid clones and respective contigs are clarified, and mutual alignment is carried out. Furthermore, the results are compared with known physical maps to map the cosmids and the contigs. In case of Corynebacterium glutamicum ATCC 13032, a physical map of Mol. Gen. Genet., 252: 255-265 (1996) can be used.

[0081] The sequence in the region which cannot be covered with the contigs (gap part) can be determined by the following method.

[0082] Clones containing sequences positioned at the ends of the contigs are selected. Among these, a clone wherein only one end of the inserted fragment has been determined is selected and the sequence at the opposite end of the inserted fragment is determined.

[0083] A shotgun library clone or a cosmid clone derived therefrom containing the sequences at the respective ends of the inserted fragments in the two contigs is identified and the full nucleotide sequence of the inserted fragment of the clone is determined.

[0084] According to this method, the nucleotide sequence of the gap part can be determined.

5

10

20

25

30

35

40

45

50

[0085] When no shotgun library clone or cosmid clone covering the gap part is available, primers complementary to the end sequences of the two different contigs are prepared and the DNA fragment in the gap part is amplified. Then, sequencing is performed by the primer walking method using the amplified DNA fragment as a template or by the shotgun method in which the sequence of a shotgun clone prepared from the amplified DNA fragment is determined. Thus, the nucleotide sequence of the above-described region can be determined.

[0086] In a region showing a low sequence accuracy, primers are synthesized using AUTOFINISH function and NAVIGATING function of consed (The University of Washington), and the sequence is determined by the primer walking method to improve the sequence accuracy.

[0087] Examples of the thus determined nucleotide sequence of the full genome include the full nucleotide sequence of genome of *Corynebacterium glutamicum* ATCC 13032 represented by SEQ ID NO:1.

(7) Determination of nucleotide sequence of microorganism genome DNA using the nucleotide sequence represented by SEQ ID NO:1

[0088] A nucleotide sequence of a polynucleotide having a homology of 80% or more with the full nucleotide sequence of Corynebacterium glutamicum ATCC 13032 represented by SEQ ID NO:1 as determined above can also be determined using the nucleotide sequence represented by SEQ ID NO:1, and the polynucleotide having a nucleotide sequence having a homology of 80% or more with the nucleotide sequence represented by SEQ ID NO:1 of the present invention is within the scope of the present invention. The term "polynucleotide having a nucleotide sequence having a homology of 80% or more with the nucleotide sequence represented by SEQ ID NO:1 of the present invention" is a polynucleotide in which a full nucleotide sequence of the chromosome DNA can be determined using as a primer an oligonucleotide composed of continuous 5 to 50 nucleotides in the nucleotide sequence represented by SEQ ID NO: 1, for example, according to PCR using the chromosome DNA as a template. A particularly preferred primer in determination of the full nucleotide sequence is an oligonucleotide having nucleotide sequences which are positioned at the interval of about 300 to 500 bp, and among such oligonucleotides, an oligonucleotide having a nucleotide sequence selected from DNAs encoding a protein relating to a main metabolic pathway is particularly preferred. The polynucleotide in which the full nucleotide sequence of the chromosome DNA can be determined using the oligonucleotide includes polynucleotides constituting a chromosome DNA derived from a microorganism belonging to coryneform bacteria. Such a polynucleotide is preferably a polynucleotide constituting chromosome DNA derived from a microorganism belonging to the genus Corynebacterium, more preferably a polynucleotide constituting a chromosome DNA of Corynebacterium glutamicum.

2. Identification of ORF (open reading frame) and expression regulatory fragment and determination of the function of ORF

[0089] Based on the full nucleotide sequence data of the genome derived from coryneform bacteria determined in the above item 1, an ORF and an expression modulating fragment can be identified. Furthermore, the function of the thus determined ORF can be determined.

[0090] The ORF means a continuous region in the nucleotide sequence of mRNA which can be translated as an amino acid sequence to mature to a protein. A region of the DNA coding for the ORF of mRNA is also called ORF.

[0091] The expression modulating fragment (hereinafter referred to as "EMF") is used herein to define a series of polynucleotide fragments which modulate the expression of the ORF or another sequence ligated operatably thereto. The expression "modulate the expression of a sequence ligated operatably" is used herein to refer to changes in the expression of a sequence due to the presence of the EMF. Examples of the EMF include a promoter, an operator, an

enhancer, a silencer, a ribosome-binding sequence, a transcriptional termination sequence, and the like. In coryneform bacteria, an EMF is usually present in an intergenic segment (a fragment positioned between two genes; about 10 to 200 nucleotides in length). Accordingly, an EMF is frequently present in an intergenic segment of 10 nucleotides or longer. It is also possible to determine or discover the presence of an EMF by using known EMF sequences as a target sequence or a target structural motif (or a target motif) using an appropriate software or comparator, such as FASTA (*Proc. Natl. Acad. Sci. USA*, 85: 2444-48 (1988)), BLAST (*J. Mol. Biol., 215:* 403-410 (1990)) or the like. Also, it can be identified and evaluated using a known EMF-capturing vector (for example, pKK232-8; manufactured by Amersham Pharmacia Biotech).

[0092] The term "target sequence" is used herein to refer to a nucleotide sequence composed of 6 or more nucleotides, an amino acid sequence composed of 2 or more amino acids, or a nucleotide sequence encoding this amino acid sequence composed of 2 or more amino acids. A longer target sequence appears at random in a data base at the lower possibility. The target sequence is preferably about 10 to 100 amino acid residues or about 30 to 300 nucleotide residues.

[0093] The term "target structural motif" or "target motif" is used herein to refer to a sequence or a combination of sequences selected optionally and reasonably. Such a motif is selected on the basis of the threedimensional structure formed by the folding of a polypeptide by means known to one of ordinary skill in the art. Various motives are known.

[0094] Examples of the target motif of a polypeptide include, but are not limited to, an enzyme activity site, a protein-protein interaction site, a signal sequence, and the like. Examples of the target motif of a nucleic acid include a promoter

protein interaction site, a signal sequence, and the like. Examples of the target motif of a nucleic acid included sequence, a transcriptional regulatory factor binding sequence, a hair pin structure, and the like.

[0095] Examples of highly useful EMF include a high-expression promoter, an inducible-expression promoter, and the like. Such an EMF can be obtained by positionally determining the nucleotide sequence of a gene which is known or expected as achieving high expression (for example, ribosomal RNA gene: GenBank Accession No. M16175 or Z46753) or a gene showing a desired induction pattern (for example, isocitrate lyase gene induced by acetic acid: Japanese Published Unexamined Patent Application No. 56782/93) via the alignment with the full genome nucleotide sequence determined in the above item 1, and isolating the genome fragment in the upstream part (usually 200 to 500 nucleotides from the translation initiation site). It is also possible to obtain a highly useful EMF by selecting an EMF showing a high expression efficiency or a desired induction pattern from among promoters captured by the EMF-capturing vector as described above.

[0096] The ORF can be identified by extracting characteristics common to individual ORFs, constructing a general model based on these characteristics, and measuring the conformity of the subject sequence with the model. In the identification, a software, such as GeneMark (*Nuc. Acids. Res., 22*: 4756-67 (1994): manufactured by GenePro)), GeneMark.hmm (manufactured by GenePro), GeneHacker (*Protein, Nucleic Acid and Enzyme, 42*: 3001-07 (1997)), Glimmer (*Nuc. Acids. Res., 26*: 544-548 (1998): manufactured by The Institute of Genomic Research), or the like, can be used. In using the software, the default (initial setting) parameters are usually used, though the parameters can be optionally changed.

[0097] In the above-described comparisons, a computer, such as UNIX, PC, Macintosh, or the like, can be used.
[0098] Examples of the ORF determined by the method of the present invention include ORFs having the nucleotide sequences represented by SEQ ID NOS:2 to 3501 present in the genome of *Corynebacterium glutamicum* as represented by SEQ ID NO:1. In these ORFs, polypeptides having the amino acid sequences represented by SEQ ID NOS:

3502 to 7001 are encoded.

5

10

15

20

25

30

35

40

50

[0099] The function of an ORF can be determined by comparing the identified amino acid sequence of the ORF with known homologous sequences using a homology searching software or comparator, such as BLAST, FAST, Smith & Waterman (*Meth. Enzym., 164*: 765 (1988)) or the like on an amino acid data base, such as Swith-Prot, PIR, GenBank-nr-aa, GenPept constituted by protein-encoding domains derived from GenBank data base, OWL or the like.

45 [0100] Furthermore, by the homology searching, the identity and similarity with the amino acid sequences of known proteins can also be analyzed.

[0101] With respect of the term "identity" used herein, where two polypeptides each having 10 amino acids are different in the positions of 3 amino acids, these polypeptides have an identity of 70% with each other. In case wherein one of the different 3 amino acids is analogue (for example, leucine and isoleucine), these polypeptides have a similarity of 80%.

[0102] As a specific example, Table 1 shows the registration numbers in known data bases of sequences which are judged as having the highest similarity with the nucleotide sequence of the ORF derived from *Corynebacterium glutamicum* ATCC 13032, genes of these sequences, functions of these genes, and identities thereof compared with known amino acid translation sequences.

[0103] Thus, a great number of novel genes derived from coryneform bacteria can be identified by determining the full nucleotide sequence of the genome derived from coryneform bacterium by the means of the present invention. Moreover, the function of the proteins encoded by these genes can be determined. Since coryneform bacteria are industrially highly useful microorganisms, many of the identified genes are industrially useful.

[0104] Moreover, the characteristics of respective microorganisms can be clarified by classifying the functions thus determined. As a result, valuable information in breeding is obtained.

[0105] Furthermore, from the ORF information derived from coryneform bacteria, the ORF corresponding to the microorganism is prepared and obtained according to the general method as disclosed in *Molecular Cloning*, 2nd ed. or the like. Specifically, an oligonucleotide having a nucleotide sequence adjacent to the ORF is synthesized, and the ORF can be isolated and obtained using the oligonucleotide as a primer and a chromosome DNA derived from coryneform bacteria as a template according to the general PCR cloning technique. Thus obtained ORF sequences include polynucleotides comprising the nucleotide sequence represented by any one of SEQ ID NOS:2 to 3501.

5

10

15

20

25

30

35

40

45

50

[0106] The ORF or primer can be prepared using a polypeptide synthesizer based on the above sequence information

[0107] Examples of the polynucleotide of the present invention include a polynucleotide containing the nucleotide sequence of the ORF obtained in the above, and a polynucleotide which hybridizes with the polynucleotide under stringent conditions.

[0108] The polynucleotide of the present invention can be a single-stranded DNA, a double-stranded DNA and a single-stranded RNA, though it is not limited thereto.

[0109] The polynucleotide which hybridizes with the polynucleotide containing the nucleotide sequence of the ORF obtained in the above under stringent conditions includes a degenerated mutant of the ORF. A degenerated mutant is a polynucleotide fragment having a nucleotide sequence which is different from the sequence of the ORF of the present invention which encodes the same amino acid sequence by degeneracy of a gene code.

[0110] Specific examples include a polynucleotide comprising the nucleotide sequence represented by any one of SEQ ID NOS:2 to 3431, and a polynucleotide which hybridizes with the polynucleotide under stringent conditions.

[0111] A polynucleotide which hybridizes under stringent conditions is a polynucleotide obtained by colony hybridization, plaque hybridization, Southern blot hybridization or the like using, as a probe, the polynucleotide having the nucleotide sequence of the ORF identified in the above. Specific examples include a polynucleotide which can be identified by carrying out hybridization at 65°C in the presence of 0.7-1.0 M NaCl using a filter on which a polynucleotide prepared from colonies or plaques is immobilized, and then washing the filter with 0.1x to 2x SSC solution (the composition of lx SSC contains 150 mM sodium chloride and 15 mM sodium citrate) at 65°C.

[0112] The hybridization can be carried out in accordance with known methods described in, for example, *Molecular Cloning*, 2nd ed., *Current Protocols in Molecular Biology, DNA Cloning 1: Core Techniques, A Practical Approach*, Second Edition, Oxford University (1995) or the like. Specific examples of the polynucleotide which can be hybridized include a DNA having a homology of 60% or more, preferably 80% or more, and particularly preferably 95% or more, with the nucleotide sequence represented by any one of SEQ ID NO:2 to 3431 when calculated using default (initial setting) parameters of a homology searching software, such as BLAST, FASTA, Smith-Waterman or the like.

[0113] Also, the polynucleotide of the present invention includes a polynucleotide encoding a polypeptide comprising the amino acid sequence represented by any one of SEQ ID NOS:3502 to 6931 and a polynucleotide which hybridizes with the polynucleotide under stringent conditions.

[0114] Furthermore, the polynucleotide of the present invention includes a polynucleotide which is present in the 5' upstream or 3' downstream region of a polynucleotide comprising the nucleotide sequence of any one of SEQ ID NOS: 2 to 3431 in a polynucleotide comprising the nucleotide sequence represented by SEQ ID NO:1, and has an activity of regulating an expression of a polypeptide encoded by the polynucleotide. Specific examples of the polynucleotide having an activity of regulating an expression of a polypeptide encoded by the polynucleotide includes a polynucleotide encoding the above described EMF, such as a promoter, an operator, an enhancer, a silencer, a ribosome-binding sequence, a transcriptional termination sequence, and the like.

[0115] The primer used for obtaining the ORF according to the above PCR cloning technique includes an oligonucleotide comprising a sequence which is the same as a sequence of 10 to 200 continuous nucleotides in the nucleotide sequence of the ORF and an adjacent region or an oligonucleotide comprising a sequence which is complementary to the oligonucleotide. Specific examples include an oligonucleotide comprising a sequence which is the same as a sequence of 10 to 200 continuous nucleotides of the nucleotide sequence represented by any one of SEQ ID NOS:1 to 3431, and an oligonucleotide comprising a sequence complementary to the oligonucleotide comprising a sequence of at least 10 to 20 continuous nucleotide of any one of SEQ ID NOS:1 to 3431. When the primers are used as a sense primer and an antisense primer, the above-described oligonucleotides in which melting temperature (T<sub>m</sub>) and the number of nucleotides are not significantly different from each other are preferred.

[0116] The oligonucleotide of the present invention includes an oligonucleotide comprising a sequence which is the same as 10 to 200 continuous nucleotides of the nucleotide sequence represented by any one of SEQ ID NOS:1 to 3431 or an oligonucleotide comprising a sequence complementary to the oligonucleotide.

[0117] Also, analogues of these oligonucleotides (hereinafter also referred to as "analogous oligonucleotides") are also provided by the present invention and are useful in the methods described herein.

[0118] Examples of the analogous oligonucleotides include analogous oligonucleotides in which a phosphodiester

bond in an oligonucleotide is converted to a phosphorothioate bond, analogous oligonucleotides in which a phosphodiester bond in an oligonucleotide is converted to an N3'-P5' phosphoamidate bond, analogous oligonucleotides in which ribose and a phosphodiester bond in an oligonucleotide is converted to a peptide nucleic acid bond, analogous oligonucleotides in which uracil in an oligonucleotide is replaced with C-5 propynyluracil, analogous oligonucleotides in which uracil in an oligonucleotide is replaced with C-5 thiazoluracil, analogous oligonucleotides in which cytosine in an oligonucleotide is replaced with C-5 propynylcytosine, analogous oligonucleotides in which cytosine in an oligonucleotide is replaced with phenoxazine-modified cytosine, analogous oligonucleotides in which ribose in an oligonucleotide is replaced with 2'-O-propylribose, analogous oligonucleotides in which ribose in an oligonucleotide with 2'-methoxyethoxyribose, and the like (*Cell Engineering*, 16: 1463 (1997)).

[0119] The above oligonucleotides and analogous oligonucleotides of the present invention can be used as probes for hybridization and antisense nucleic acids described below in addition to as primers.

**[0120]** Examples of a primer for the antisense nucleic acid techniques known in the art include an oligonucleotide which hybridizes the oligonucleotide of the present invention under stringent conditions and has an activity regulating expression of the polypeptide encoded by the polynucleotide, in addition to the above oligonucleotide.

### 3. Determination of isozymes

[0121] Many mutants of coryneform bacteria which are useful in the production of useful substances, such as amino acids, nucleic acids, vitamins, saccharides, organic acids, and the like, are obtained by the present invention.

[0122] However, since the gene sequence data of the microorganism has been, to date, insufficient, useful mutants have been obtained by mutagenic techniques using a mutagen, such as nitrosoguanidine (NTG) or the like.

[0123] Although genes can be mutated randomly by the mutagenic method using the above-described mutagen, all genes encoding respective isozymes having similar properties relating to the metabolism of intermediates cannot be mutated. In the mutagenic method using a mutagen, genes are mutated randomly. Accordingly, harmful mutations worsening culture characteristics, such as delay in growth, accelerated foaming, and the like, might be imparted at a great frequency, in a random manner.

[0124] However, if gene sequence information is available, such as is provided by the present invention, it is possible to mutate all of the genes encoding target isozymes. In this case, harmful mutations may be avoided and the target mutation can be incorporated.

[0125] Namely, an accurate number and sequence information of the target isozymes in coryneform bacteria can be obtained based on the ORF data obtained in the above item 2. By using the sequence information, all of the target isozyme genes can be mutated into genes having the desired properties by, for example, the site-specific mutagenesis method described in *Molecular Cloning*, 2nd ed. to obtain useful mutants having elevated productivity of useful substances.

4. Clarification or determination of biosynthesis pathway and signal transmission pathway

[0126] Attempts have been made to elucidate biosynthesis pathways and signal transmission pathways in a number of organisms, and many findings have been reported. However, there are many unknown aspects of coryneform bacteria since a number of genes have not been identified so far.

[0127] These unknown points can be clarified by the following method.

[0128] The functional information of ORF derived from coryneform bacteria as identified by the method of above item 2 is arranged. The term "arranged" means that the ORF is classified based on the biosynthesis pathway of a substance or the signal transmission pathway to which the ORF belongs using known information according to the functional information. Next, the arranged ORF sequence information is compared with enzymes on the biosynthesis pathways or signal transmission pathways of other known organisms. The resulting information is combined with known data on coryneform bacteria. Thus, the biosynthesis pathways and signal transmission pathways in coryneform bacteria, which have been unknown so far, can be determined.

[0129] As a result that these pathways which have been unknown or unclear hitherto are clarified, a useful mutant for producing a target useful substance can be efficiently obtained.

[0130] When the thus clarified pathway is judged as important in the synthesis of a useful product, a useful mutant can be obtained by selecting a mutant wherein this pathway has been strengthened. Also, when the thus clarified pathway is judged as not important in the biosynthesis of the target useful product, a useful mutant can be obtained by selecting a mutant wherein the utilization frequency of this pathway is lowered.

5. Clarification or determination of useful mutation point

[0131] Many useful mutants of coryneform bacteria which are suitable for the production of useful substances, such

16

15

20

25

10

30

35

45

50

55

as amino acids, nucleic acids, vitamins, saccharides, organic acids, and the like, have been obtained. However, it is hardly known which mutation point is imparted to a gene to improve the productivity.

[0132] However, mutation points contained in production strains can be identified by comparing desired sequences of the genome DNA of the production strains obtained from coryneform bacteria by the mutagenic technique with the nucleotide sequences of the corresponding genome DNA and ORF derived from coryneform bacteria determined by the methods of the above items 1 and 2 and analyzing them

[0133] Moreover, effective mutation points contributing to the production can be easily specified from among these mutation points on the basis of known information relating to the metabolic pathways, the metabolic regulatory mechanisms, the structure activity correlation of enzymes, and the like.

[0134] When any efficient mutation can be hardly specified based on known data, the mutation points thus identified can be introduced into a wild strain of coryneform bacteria or a production strain free of the mutation. Then, it is examined whether or not any positive effect can be achieved on the production.

[0135] For example, by comparing the nucleotide sequence of homoserine dehydrogenase gene *hom* of a lysine-producing B-6 strain of *Corynebacterium glutamicum* (*Appl. Microbiol. Biotechnol., 32*: 269-273 (1989)) with the nucleotide sequence corresponding to the genome of *Corynebacterium glutamicum* ATCC 13032 according to the present invention, a mutation of amino acid replacement in which valine at the 59-position is replaced with alanine (Val59Ala) was identified. A strain obtained by introducing this mutation into the ATCC 13032 strain by the gene replacement method can produce lysine, which indicates that this mutation is an effective mutation contributing to the production of lysine.

[0136] Similarly, by comparing the nucleotide sequence of pyruvate carboxylase gene *pyc* of the B-6 strain with the nucleotide sequence corresponding to the ATCC 13032 genome, a mutation of amino acid replacement in which proline at the 458-position was replaced with serine (Pro458Ser) was identified. A strain obtained by introducing this mutation into a lysine-producing strain of No. 58 (FERM BP-7134) of *Corynebacterium glutamicum* free of this mutation shows an improved lysine productivity in comparison with the No. 58 strain, which indicates that this mutation is an effective mutation contributing to the production of lysine.

[0137] In addition, a mutation A1a213Thr in glucose-6-phosphate dehydrogenase was specified as an effective mutation relating to the production of lysine by detecting glucose-6-phosphate dehydrogenase gene *zwf* of the B-6 strain.
[0138] Furthermore, the lysine-productivity of *Corynebacterium glutamicum* was improved by replacing the base at the 932-position of aspartokinase gene *lysC* of the *Corynebacterium glutamicum* ATCC 13032 genome with cytosine to thereby replace threonine at the 311-position by isoleucine, which indicates that this mutation is an effective mutation contributing to the production of lysine.

[0139] Also, as another method to examine whether or not the identified mutation point is an effective mutation, there is a method in which the mutation possessed by the lysine-producing strain is returned to the sequence of a wild type strain by the gene replacement method and whether or not it has a negative influence on the lysine productivity. For example, when the amino acid replacement mutation Val59Ala possessed by *hom* of the lysine-producing B-6 strain was returned to a wild type amino acid sequence, the lysine productivity was lowered in comparison with the B-6 strain. Thus, it was found that this mutation is an effective mutation contributing to the production of lysine.

[0140] Effective mutation points can be more efficiently and comprehensively extracted by combining, if needed, the DNA array analysis or proteome analysis described below.

6. Method of breeding industrially advantageous production strain

5

10

15

20

25

30

35

40

45

50

55

[0141] It has been a general practice to construct production strains, which are used industrially in the fermentation production of the target useful substances, such as amino acids, nucleic acids, vitamins, saccharides, organic acids, and the like, by repeating mutagenesis and breeding based on random mutagenesis using mutagens, such as NTG or the like, and screening.

[0142] In recent years, many examples of improved production strains have been made through the use of recombinant DNA techniques. In breeding, however, most of the parent production strains to be improved are mutants obtained by a conventional mutagenic procedure (W. Leuchtenberger, *Amino Acids - Technical Production and Use.* In: Roehr (ed) Biotechnology, second edition, vol. 6, products of primary metabolism. VCH Verlagsgesellschaft mbH, Weinheim, P 465 (1996)).

[0143] Although mutagenesis methods have largely contributed to the progress of the fermentation industry, they suffer from a serious problem of multiple, random introduction of mutations into every part of the chromosome. Since many mutations are accumulated in a single chromosome each time a strain is improved, a production strain obtained by the random mutation and selecting is generally inferior in properties (for example, showing poor growth, delayed consumption of saccharides, and poor resistance to stresses such as temperature and oxygen) to a wild type strain, which brings about troubles such as failing to establish a sufficiently elevated productivity, being frequently contaminated with miscellaneous bacteria, requiring troublesome procedures in culture maintenance, and the like, and, in its

turn, elevating the production cost in practice. In addition, the improvement in the productivity is based on random mutations and thus the mechanism thereof is unclear. Therefore, it is very difficult to plan a rational breeding strategy for the subsequent improvement in the productivity.

[0144] According to the present invention, effective mutation points contributing to the production can be efficiently specified from among many mutation points accumulated in the chromosome of a production strain which has been bred from coryneform bacteria and, therefore, a novel breeding method of assembling these effective mutations in the coryneform bacteria can be established. Thus, a useful production strain can be reconstructed. It is also possible to construct a useful production strain from a wild type strain.

[0145] Specifically, a useful mutant can be constructed in the following manner.

5

10

15

20

25

30

35

40

45

50

55

[0146] One of the mutation points is incorporated into a wild type strain of coryneform bacteria. Then, it is examined whether or not a positive effect is established on the production. When a positive effect is obtained, the mutation point is saved. When no effect is obtained, the mutation point is removed. Subsequently, only a strain having the effective mutation point is used as the parent strain, and the same procedure is repeated. In general, the effectiveness of a mutation positioned upstream cannot be clearly evaluated in some cases when there is a rate-determining point in the downstream of a biosynthesis pathway. It is therefore preferred to successively evaluate mutation points upward from downstream.

[0147] By reconstituting effective mutations by the method as described above in a wild type strain or a strain which has a high growth speed or the same ability to consume saccharides as the wild type strain, it is possible to construct an industrially advantageous strain which is free of troubles in the previous methods as described above and to conduct fermentation production using such strains within a short time or at a higher temperature.

[0148] For example, a lysine-producing mutant B-6 (*Appl. Microbiol. Biotechnol., 32*: 262-273 (1989)), which is obtained by multiple rounds of random mutagenesis from a wild type strain *Corynebacterium glutamicum* ATCC 13032, enables lysine fermentation to be performed at a temperature between 30 and 34°C but shows lowered growth and lysine productivity at a temperature exceeding 34°C. Therefore, the fermentation temperature should be maintained at 34°C or lower. In contrast thereto, the production strain described in the above item 5, which is obtained by reconstituting effective mutations relating to lysine production, can achieve a productivity at 40 to 42°C equal or superior to the result obtained by culturing at 30 to 34°C. Therefore, this strain is industrially advantageous since it can save the load of cooling during the fermentation.

[0149] When culture should be carried out at a high temperature exceeding 43°C, a production strain capable of conducting fermentation production at a high temperature exceeding 43°C can be obtained by reconstituting useful mutations in a microorganism belonging to the genus *Corynebacterium* which can grow at high temperature exceeding 43°C. Examples of the microorganism capable of growing at a high temperature exceeding 43°C include *Corynebacterium thermoaminogenes*, such as *Corynebacterium thermoaminogenes* FERM 9244, FERM 9245, FERM 9246 and FERM 9247.

[0150] A strain having a further improved productivity of the target product can be obtained using the thus reconstructed strain as the parent strain and further breeding it using the conventional mutagenesis method, the gene amplification method, the gene replacement method using the recombinant DNA technique, the transduction method or the cell fusion method. Accordingly, the microorganism of the present invention includes, but is not limited to, a mutant, a cell fusion strain, a transformant, a transductant or a recombinant strain constructed by using recombinant DNA techniques, so long as it is a producing strain obtained via the step of accumulating at least two effective mutations in a coryneform bacteria in the course of breeding.

[0151] When a mutation point judged as being harmful to the growth or production is specified, on the other hand, it is examined whether or not the producing strain used at present contains the mutation point. When it has the mutation, it can be returned to the wild type gene and thus a further useful production strain can be bred.

[0152] The breeding method as described above is applicable to microorganisms, other than coryneform bacteria, which have industrially advantageous properties (for example, microorganisms capable of quickly utilizing less expensive carbon sources, microorganisms capable of growing at higher temperatures).

- 7. Production and utilization of polynucleotide array
- (1) Production of polynucleotide array

[0153] A polynucleotide array can be produced using the polynucleotide or oligonucleotide of the present invention obtained in the above items 1 and 2.

[0154] Examples include a polynucleotide array comprising a solid support to which at least one of a polynucleotide comprising the nucleotide sequence represented by SEQ ID NOS:2 to 3501, a polynucleotide which hybridizes with the polynucleotide under stringent conditions, and a polynucleotide comprising 10 to 200 continuous nucleotides in the nucleotide sequence of the polynucleotide is adhered; and a polynucleotide array comprising a solid support to

which at least one of a polynucleotide encoding a polypeptide comprising the amino acid sequence represented by any one of SEQ ID NOS:3502 to 7001, a polynucleotide which hybridizes with the polynucleotide under stringent conditions, and a polynucleotide comprising 10 to 200 continuous bases in the nucleotide sequences of the polynucleotides is adhered.

[0155] Polynucleotide arrays of the present invention include substrates known in the art, such as a DNA chip, a DNA microarray and a DNA macroarray, and the like, and comprises a solid support and plural polynucleotides or fragments thereof which are adhered to the surface of the solid support.

[0156] Examples of the solid support include a glass plate, a nylon membrane, and the like.

[0157] The polynucleotides or fragments thereof adhered to the surface of the solid support can be adhered to the surface of the solid support using the general technique for preparing arrays. Namely, a method in which they are adhered to a chemically surface-treated solid support, for example, to which a polycation such as polylysine or the like has been adhered (*Nat. Genet.*, 21: 15-19 (1999)). The chemically surface-treated supports are commercially available and the commercially available solid product can be used as the solid support of the polynucleotide array according to the present invention.

[0158] As the polynucleotides or oligonucleotides adhered to the solid support, the polynucleotides and oligonucleotides of the present invention obtained in the above items 1 and 2 can be used.

[0159] The analysis described below can be efficiently performed by adhering the polynucleotides or oligonucleotides to the solid support at a high density, though a high fixation density is not always necessary.

[0160] Apparatus for achieving a high fixation density, such as an arrayer robot or the like, is commercially available from Takara Shuzo (GMS417 Arrayer), and the commercially available product can be used.

[0161] Also, the oligonucleotides of the present invention can be synthesized directly on the solid support by the photolithography method or the like (*Nat. Genet., 21*: 20-24 (1999)). In this method, a linker having a protective group which can be removed by light irradiation is first adhered to a solid support, such as a slide glass or the like. Then, it is irradiated with light through a mask (a photolithograph mask) permeating light exclusively at a definite part of the adhesion part. Next, an oligonucleotide having a protective group which can be removed by light irradiation is added to the part. Thus, a ligation reaction with the nucleotide arises exclusively at the irradiated part. By repeating this procedure, oligonucleotides, each having a desired sequence, different from each other can be synthesized in respective parts. Usually, the oligonucleotides to be synthesized have a length of 10 to 30 nucleotides.

30 (2) Use of polynucleotide array

5

10

15

20

25

40

45

50

55

[0162] The following procedures (a) and (b) can be carried out using the polynucleotide array prepared in the above (1).

(a) Identification of mutation point of coryneform bacterium mutant and analysis of expression amount and expression profile of gene encoded by genome

[0163] By subjecting a gene derived from a mutant of coryneform bacteria or an examined gene to the following steps (i) to (iv), the mutation point of the gene can be identified or the expression amount and expression profile of the gene can be analyzed:

- (i) producing a polynucleotide array by the method of the above (1);
- (ii) incubating polynucleotides immobilized on the polynucleotide array together with the labeled gene derived from a mutant of the coryneform bacterium using the polynucleotide array produced in the above (i) under hybridization conditions;
- (iii) detecting the hybridization; and
- (iv) analyzing the hybridization data.

[0164] The gene derived from a mutant of coryneform bacteria or the examined gene include a gene relating to biosynthesis of at least one selected from amino acids, nucleic acids, vitamins, saccharides, organic acids, and analogues thereof

[0165] The method will be described in detail.

[0166] A single nucleotide polymorphism (SNP) in a human region of 2,300 kb has been identified using polynucleotide arrays (*Science, 280*: 1077-82 (1998)). In accordance with the method of identifying SNP and methods described in *Science, 278*: 680-686 (1997); *Proc. Natl. Acad. Sci. USA, 96*: 12833-38 (1999); *Science, 284*: 1520-23 (1999), and the like using the polynucleotide array produced in the above (1) and a nucleic acid molecule (DNA, RNA) derived from coryneform bacteria in the method of the hybridization, a mutation point of a useful mutant, which is useful in producing an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, or the like can be identified and the gene

expression amount and the expression profile thereof can be analyzed.

10

25

35

50

55

[0167] The nucleic acid molecule (DNA, RNA) derived from the coryneform bacteria can be obtained according to the general method described in *Molecular Cloning*, 2nd ed. or the like. mRNA derived from *Corynebacterium glutamicum* can also be obtained by the method of Bormann et al. (*Molecular Microbiology*, 6: 317-326 (1992)) or the like.

- [0168] Although ribosomal RNA (rRNA) is usually obtained in large excess in addition to the target mRNA, the analysis is not seriously disturbed thereby.
  - [0169] The resulting nucleic acid molecule derived from coryneform bacteria is labeled. Labeling can be carried out according to a method using a fluorescent dye, a method using a radioisotope or the like.
  - [0170] Specific examples include a labeling method in which psoralen-biotin is crosslinked with RNA extracted from a microorganism and, after hybridization reaction, a fluorescent dye having streptoavidin bound thereto is bound to the biotin moiety (*Nat. Biotechnol., 16*: 45-48 (1998)); a labeling method in which a reverse transcription reaction is carried out using RNA extracted from a microorganism as a template and random primers as primers, and dUTP having a fluorescent dye (for example, Cy3, Cy5) (manufactured by Amersham Pharmacia Biotech) is incorporated into cDNA (*Proc. Natl. Acad. Sci. USA, 96.* 12833-38 (1999)); and the like.
- 15 [0171] The labeling specificity can be improved by replacing the random primers by sequences complementary to the 3'-end of ORF (*J. Bacteriol., 181*: 6425-40 (1999)).
  - [0172] In the hybridization method, the hybridization and subsequent washing can be carried out by the general method (*Nat. Bioctechnol.*, 14: 1675-80 (1996), or the like).
- [0173] Subsequently, the hybridization intensity is measured depending on the hybridization amount of the nucleic acid molecule used in the labeling. Thus, the mutation point can be identified and the expression amount of the gene can be calculated.
  - [0174] The hybridization intensity can be measured by visualizing the fluorescent signal, radioactivity, luminescence dose, and the like, using a laser confocal microscope, a CCD camera, a radiation imaging device (for example, STORM manufactured by Amersham Pharmacia Biotech), and the like, and then quantifying the thus visualized data.
  - [0175] A polynucleotide array on a solid support can also be analyzed and quantified using a commercially available apparatus, such as GMS418 Array Scanner (manufactured by Takara Shuzo) or the like.
    - [0176] The gene expression amount can be analyzed using a commercially available software (for example, ImaGene manufactured by Takara Shuzo; Array Gauge manufactured by Fuji Photo Film; ImageQuant manufactured by Amersham Pharmacia Biotech, or the like).
- [0177] A fluctuation in the expression amount of a specific gene can be monitored using a nucleic acid molecule obtained in the time course of culture as the nucleic acid molecule derived from coryneform bacteria. The culture conditions can be optimized by analyzing the fluctuation.
  - [0178] The expression profile of the microorganism at the total gene level (namely, which genes among a great number of genes encoded by the genome have been expressed and the expression ratio thereof) can be determined using a nucleic acid molecule having the sequences of many genes determined from the full genome sequence of the microorganism. Thus, the expression amount of the genes determined by the full genome sequence can be analyzed and, in its turn, the biological conditions of the microorganism can be recognized as the expression pattern at the full gene level.
- 40 (b) Confirmation of the presence of gene homologous to examined gene in coryneform bacteria
  - [0179] Whether or not a gene homologous to the examined gene, which is present in an organism other than coryneform bacteria, is present in coryneform bacteria can be detected using the polynucleotide array prepared in the above (1).
- [0180] This detection can be carried out by a method in which an examined gene which is present in an organism other than coryneform bacteria is used instead of the nucleic acid molecule derived from coryneform bacteria used in the above identification/analysis method of (1).
  - 8. Recording medium storing full genome nucleotide sequence and ORF data and being readable by a computer and methods for using the same
  - [0181] The term "recording medium or storage device which is readable by a computer" means a recording medium or storage medium which can be directly readout and accessed with a computer. Examples include magnetic recording media, such as a floppy disk, a hard disk, a magnetic tape, and the like; optical recording media, such as CD-ROM, CD-R, CD-RW, DVD-ROM, DVD-RAM, DVD-RW, and the like; electric recording media, such as RAM, ROM, and the like; and hybrids in these categories (for example, magnetic/optical recording media, such as MO and the like).
  - [0182] Instruments for recording or inputting in or on the recording medium or instruments or devices for reading out the information in the recording medium can be appropriately selected, depending on the type of the recording medium

and the access device utilized. Also, various data processing programs, software, comparator and formats are used for recording and utilizing the polynucleotide sequence information or the like, of the present invention in the recording medium. The information can be expressed in the form of a binary file, a text file or an ASCII file formatted with commercially available software, for example. Moreover, software for accessing the sequence information is available and known to one of ordinary skill in the art.

5

10

15

20

25

40

45

50

55

[0183] Examples of the information to be recorded in the above-described medium include the full genome nucleotide sequence information of coryneform bacteria as obtained in the above item 2, the nucleotide sequence information of ORF, the amino acid sequence information encoded by the ORF, and the functional information of polynucleotides coding for the amino acid sequences.

[0184] The recording medium or storage device which is readable by a computer according to the present invention refers to a medium in which the information of the present invention has been recorded. Examples include recording media or storage devices which are readable by a computer storing the nucleotide sequence information represented by SEQ ID NOS:1 to 3501, the amino acid sequence information represented by SEQ ID NOS:3502 to 7001, the functional information of the nucleotide sequences represented by SEQ ID NOS:1 to 3501, the functional information of the amino acid sequences represented by SEQ ID NOS:3502 to 7001, and the information listed in Table 1 below and the like.

- 9. System based on a computer using the recording medium of the present invention which is readable by a computer
- [0185] The term "system based on a computer" as used herein refers a system composed of hardware device(s), software device(s), and data recording device(s) which are used for analyzing the data recorded in the recording medium of the present invention which is readable by a computer.
  - [0186] The hardware device(s) are, for example, composed of an input unit, a data recording unit, a central processing unit and an output unit collectively or individually.
- [0187] By the software device(s), the data recorded in the recording medium of the present invention are searched or analyzed using the recorded data and the hardware device(s) as described herein. Specifically, the software device (s) contain at least one program which acts on or with the system in order to screen, analyze or compare biologically meaningful structures or information from the nucleotide sequences, amino acid sequences and the like recorded in the recording medium according to the present invention.
- [0188] Examples of the software device(s) for identifying ORF and EMF domains include GeneMark (*Nuc. Acids. Res., 22*: 4756-67 (1994)), GeneHacker (*Protein, Nucleic Acid and Enzyme, 42*: 3001-07 (1997)), Glimmer (The Institute of Genomic Research; *Nuc. Acids. Res., 26*: 544-548 (1998)) and the like. In the process of using such a software device, the default (initial setting) parameters are usually used, although the parameters can be changed, if necessary, in a manner known to one of ordinary skill in the art.
- [0189] Examples of the software device(s) for identifying a genome domain or a polypeptide domain analogous to the target sequence or the target structural motif (homology searching) include FASTA, BLAST, Smith-Waterman, GenetyxMac (manufactured by Software Development), GCG Package (manufactured by Genetic Computer Group), GenCore (manufactured by Compugen), and the like. In the process of using such a software device, the default (initial setting) parameters are usually used, although the parameters can be changed, if necessary, in a manner known to one of ordinary skill in the art.
- [0190] Such a recording medium storing the full genome sequence data is useful in preparing a polynucleotide array by which the expression amount of a gene encoded by the genome DNA of coryneform bacteria and the expression profile at the total gene level of the microorganism, namely, which genes among many genes encoded by the genome have been expressed and the expression ratio thereof, can be determined.
- [0191] The data recording device(s) provided by the present invention are, for example, memory device(s) for recording the data recorded in the recording medium of the present invention and target sequence or target structural motif data, or the like, and a memory accessing device(s) for accessing the same.
  - [0192] Namely, the system based on a computer according to the present invention comprises the following:
    - (i) a user input device that inputs the information stored in the recording medium of the present invention, and target sequence or target structure motif information;
    - (ii) a data storage device for at least temporarily storing the input information;
    - (iii) a comparator that compares the information stored in the recording medium of the present invention with the target sequence or target structure motif information, recorded by the data storing device of (ii) for screening and analyzing nucleotide sequence information which is coincident with or analogous to the target sequence or target structure motif information; and
    - (iv) an output device that shows a screening or analyzing result obtained by the comparator.

[0193] This system is usable in the methods in items 2 to 5 as described above for searching and analyzing the ORF and EMF domains, target sequence, target structural motif, etc. of a coryneform bacterium, searching homologs, searching and analyzing isozymes, determining the biosynthesis pathway and the signal transmission pathway, and identifying spots which have been found in the proteome analysis. The term "homologs" as used herein includes both of orthologs and paralogs.

10. Production of polypeptide using ORF derived from coryneform bacteria

10

15

20

25

35

40

45

50

55

[0194] The polypeptide of the present invention can be produced using a polynucleotide comprising the ORF obtained in the above item 2. Specifically, the polypeptide of the present invention can be produced by expressing the polynucleotide of the present invention or a fragment thereof in a host cell, using the method described in *Molecular Cloning*, 2nd ed., *Current Protocols in Molecular Biology*, and the like, for example, according to the following method.

[0195] A DNA fragment having a suitable length containing a part encoding the polypeptide is prepared from the full length ORF sequence, if necessary.

[0196] Also, DNA in which nucleotides in a nucleotide sequence at a part encoding the polypeptide of the present invention are replaced to give a codon suitable for expression of the host cell, if necessary. The DNA is useful for efficiently producing the polypeptide of the present invention.

[0197] A recombinant vector is prepared by inserting the DNA fragment into the downstream of a promoter in a suitable expression vector.

[0198] The recombinant vector is introduced to a host cell suitable for the expression vector.

[0199] Any of bacteria, yeasts, animal cells, insect cells, plant cells, and the like can be used as the host cell so long as it can be expressed in the gene of interest.

[0200] Examples of the expression vector include those which can replicate autonomously in the above-described host cell or can be integrated into chromosome and have a promoter at such a position that the DNA encoding the polypeptide of the present invention can be transcribed.

[0201] When a procaryote cell, such as a bacterium or the like, is used as the host cell, it is preferred that the recombinant vector containing the DNA encoding the polypeptide of the present invention can replicate autonomously in the bacterium and is a recombinant vector constituted by, at least a promoter, a ribosome binding sequence, the DNA of the present invention and a transcription termination sequence. A promoter controlling gene can also be contained therewith in operable combination.

[0202] Examples of the expression vectors include a vector plasmid which is replicable in Corynebacterium glutamicum, such as pCGI (Japanese Published Unexamined Patent Application No. 134500/82), pCG2 (Japanese Published Unexamined Patent Application No. 35197/83), pCG4 (Japanese Published Unexamined Patent Application No. 183799/82), pCG11 (Japanese Published Unexamined Patent Application No. 134500/82), pCG116, pCE54 and pCB101 (Japanese Published Unexamined Patent Application No. 105999/83), pCE51, pCE52 and pCE53 (Mol. Gen. Genet., 196: 175-178 (1984)), and the like; a vector plasmid which is replicable in Escherichia coli, such as pET3 and pET11 (manufactured by Stratagene), pBAD, pThioHis and pTrcHis (manufactured by Invitrogen), pKK223-3 and pGEX2T (manufactured by Amersham Pharmacia Biotech), and the like; and pBTrp2, pBTac1 and pBTac2 (manufactured by Boehringer Mannheim Co.), pSE280 (manufactured by Invitrogen), pGEMEX-1 (manufactured by Promega), pQE-8 (manufactured by QIAGEN), pKYP10 (Japanese Published Unexamined Patent Application No. 110600/83), pKYP200 (Agric. Biol. Chem., 48: 669 (1984)), pLSA1 (Agric. Biol. Chem., 53: 277 (1989)), pGEL1 (Proc. Natl. Acad. Sci. USA, 82: 4306 (1985)), pBluescript II SK(-) (manufactured by Stratagene), pTrs30 (prepared from Escherichia coli JM109/pTrS30 (FERM BP-5407)), pTrs32 (prepared from Escherichia coli JM109/pTrS32 (FERM BP-5408)), pGHA2 (prepared from Escherichia coli IGHA2 (FERM B-400), Japanese Published Unexamined Patent Application No. 221091/85), pGKA2 (prepared from Escherichia coli IGKA2 (FERM BP-6798), Japanese Published Unexamined Patent Application No. 221091/85), pTerm2 (U.S. Patents 4,686,191, 4,939,094 and 5,160,735), pSupex, pUB110, pTP5, pC194 and pEG400 (J. Bacteriol., 172: 2392 (1990)), pGEX (manufactured by Pharmacia), pET system (manufactured by Novagen), and the like.

[0203] Any promoter can be used so long as it can function in the host cell. Examples include promoters derived from *Escherichia coli*, phage and the like, such as *trp* promoter ( $P_{trp}$ ), *lac* promoter,  $P_L$  promoter,  $P_R$  promoter, T7 promoter and the like. Also, artificially designed and modified promoters, such as a promoter in which two  $P_{trp}$  are linked in series ( $P_{trp} \times 2$ ), *tac* promoter, *lac*T7 promoter *let*l promoter and the like, can be used.

[0204] It is preferred to use a plasmid in which the space between Shine-Dalgarno sequence which is the ribosome binding sequence and the initiation codon is adjusted to an appropriate distance (for example, 6 to 18 nucleotides).

[0205] The transcription termination sequence is not always necessary for the expression of the DNA of the present invention. However, it is preferred to arrange the transcription terminating sequence at just downstream of the structural gene.

[0206] One of ordinary skill in the art will appreciate that the codons of the above-described elements may be opti-

mized, in a known manner, depending on the host cells and environmental conditions utilized.

5

10

15

20

30

35

40

55

[0207] Examples of the host cell include microorganisms belonging to the genus Escherichia, the genus Serratia, the genus Bacillus, the genus Brevibacterium, the genus Corynebacterium, the genus Microbacterium, the genus Pseudomonas, and the like. Specific examples include Escherichia coli XL1-Blue, Escherichia coli XL2-Blue, Escherichia coli XL2-Blue, Escherichia coli DH1, Escherichia coli MC1000, Escherichia coli KY3276, Escherichia coli W1485, Escherichia coli JM109, Escherichia coli HB101, Escherichia coli No. 49, Escherichia coli W3110, Escherichia coli NY49, Escherichia coli GI698, Escherichia coli TB1, Serratia ficaria, Serratia fonticola, Serratia liquefaciens, Serratia marcescens, Bacillus subtilis, Bacillus amyloliquefaciens, Corynebacterium ammonia genes, Brevibacterium immariophilum ATCC 14068, Brevibacterium saccharolyticum ATCC 14066, Corynebacterium glutamicum ATCC 13032, Corynebacterium glutamicum ATCC 13869, Corynebacterium glutamicum ATCC 14067 (prior genus and species: Brevibacterium flavum), Corynebacterium lactofermentum), Corynebacterium acetoacidophilum ATCC 13870, Corynebacterium thermoaminogenes FERM 9244, Microbacterium ammoniaphilum ATCC 15354, Pseudomonas putida, Pseudomonas sp. D-0110, and the like.

[0208] When Corynebacterium glutamicum or an analogous microorganism is used as a host, an EMF necessary for expressing the polypeptide is not always contained in the vector so long as the polynucleotide of the present invention contains an EMF. When the EMF is not contained in the polynucleotide, it is necessary to prepare the EMF separately and ligate it so as to be in operable combination. Also, when a higher expression amount or specific expression regulation is necessary, it is necessary to ligate the EMF corresponding thereto so as to put the EMF in operable combination with the polynucleotide. Examples of using an externally ligated EMF are disclosed in Microbiology, 142: 1297-1309 (1996).

[0209] With regard to the method for the introduction of the recombinant vector, any method for introducing DNA into the above-described host cells, such as a method in which a calcium ion is used (*Proc. Natl. Acad. Sci. USA, 69*: 2110 (1972)), a protoplast method (Japanese Published Unexamined Patent Application No. 2483942/88), the methods described in *Gene, 17*: 107 (1982) and *Molecular & General Genetics, 168*: 111 (1979) and the like, can be used.

[0210] When yeast is used as the host cell, examples of the expression vector include pYES2 (manufactured by Invitrogen), YEp13 (ATCC 37115), YEp24 (ATCC 37051), YCp50 (ATCC 37419), pHS19, pHS15, and the like.

[0211] Any promoter can be used so long as it can be expressed in yeast. Examples include a promoter of a gene in the glycolytic pathway, such as hexose kinase and the like, PHO5 promoter, PGK promoter, GAP promoter, ADH promoter, gal 1 promoter, gal 10 promoter, a heat shock protein promoter, MF all promoter, CUP 1 promoter, and the like. [0212] Examples of the host cell include microorganisms belonging to the genus Saccharomyces, the genus Schizosaccharomyces, the genus Kluyveromyces, the genus Trichosporon, the genus Schwanniomyces, the genus

Pichia, the genus Candida and the like. Specific examples include Saccharomyces cerevisiae, Schizosaccharomyces pombe, Kluyveromyces lactis, Trichosporon pullulans, Schwanniomyces alluvius, Candida utilis and the like.

[0213] With regard to the method for the introduction of the recombinant vector, any method for introducing DNA into yeast, such as an electroporation method (*Methods. Enzymol., 194*: 182 (1990)), a spheroplast method (*Proc. Natl. Acad. Sci. USA, 75*: 1929 (1978)), a lithium acetate method (*J. Bacteriol., 153*: 163 (1983)), a method described in *Proc. Natl. Acad. Sci. USA, 75*: 1929 (1978) and the like, can be used.

[0214] When animal cells are used as the host cells, examples of the expression vector include pcDNA3.1, pSinRep5 and pCEP4 (manufactured by Invitorogen), pRev-Tre (manufactured by Clontech), pAxCAwt (manufactured by Takara Shuzo), pcDNAI and pcDM8 (manufactured by Funakoshi), pAGE107 (Japanese Published Unexamined Patent Application No. 22979/91; *Cytotechnology, 3*:133 (1990)), pAS3-3 (Japanese Published Unexamined Patent Application No. 227075/90), pcDM8 (*Nature, 329*: 840 (1987)), pcDNAI/Amp (manufactured by Invitrogen), pREP4 (manufactured by Invitrogen), pAGE103 (*J. Biochem., 101*: 1307 (1987)), pAGE210, and the like.

[0215] Any promoter can be used so long as it can function in animal cells. Examples include a promoter of IE-(immediate early) gene of cytomegalovirus (CMV), an early promoter of SV40, a promoter of retrovirus, a metal-lothionein promoter, a heat shock promoter, SRα promoter, and the like. Also, the enhancer of the IE gene of human CMV can be used together with the promoter.

[0216] Examples of the host cell include human Namalwa cell, monkey COS cell, Chinese hamster CHO cell, HST5637 (Japanese Published Unexamined Patent Application No. 299/88), and the like.

[0217] The method for introduction of the recombinant vector into animal cells is not particularly limited, so long as it is the general method for introducing DNA into animal cells, such as an electroporation method (*Cytotechnology, 3*: 133 (1990)), a calcium phosphate method (Japanese Published Unexamined Patent Application No. 227075/90), a lipofection method (*Proc. Natl. Acad. Sci. USA, 84*, 7413 (1987)), the method described in *Virology, 52*: 456 (1973), and the like.

[0218] When insect cells are used as the host cells, the polypeptide can be expressed, for example, by the method described in *Bacurovirus Expression Vectors*, *A Laboratory Manual*, W.H. Freeman and Company, New York (1992), *Bio/Technology*, 6: 47 (1988), or the like.

[0219] Specifically, a recombinant gene transfer vector and bacurovirus are simultaneously inserted into insect cells

to obtain a recombinant virus in an insect cell culture supernatant, and then the insect cells are infected with the resulting recombinant virus to express the polypeptide.

[0220] Examples of the gene introducing vector used in the method include pBlueBac4.5, pVL1392, pVL1393 and pBlueBacIII (manufactured by Invitrogen), and the like.

- [0221] Examples of the bacurovirus include Autographa californica nuclear polyhedrosis virus with which insects of the family *Barathra* are infected, and the like.
  - [0222] Examples of the insect cells include *Spodoptera frugiperda* oocytes Sf9 and Sf21 (*Bacurovirus Expression Vectors, A Laboratory Manual,* W.H. Freeman and Company, New York (1992)), *Trichoplusia ni* oocyte High 5 (manufactured by Invitrogen) and the like.
- 10 [0223] The method for simultaneously incorporating the above-described recombinant gene transfer vector and the above-described bacurovirus for the preparation of the recombinant virus include calcium phosphate method (Japanese Published Unexamined Patent Application No. 227075/90), lipofection method (*Proc. Natl. Acad. Sci. USA, 84*: 7413 (1987)) and the like.
  - [0224] When plant cells are used as the host cells, examples of expression vector include a Ti plasmid, a tobacco mosaic virus vector, and the like.
  - [0225] Any promoter can be used so long as it can be expressed in plant cells. Examples include 35S promoter of cauliflower mosaic virus (CaMV), rice actin 1 promoter, and the like.
  - [0226] Examples of the host cells include plant cells and the like, such as tobacco, potato, tomato, carrot, soybean, rape, alfalfa, rice, wheat, barley, and the like.
  - [0227] The method for introducing the recombinant vector is not particularly limited, so long as it is the general method for introducing DNA into plant cells, such as the *Agrobacterium* method (Japanese Published Unexamined Patent Application No. 140885/84, Japanese Published Unexamined Patent Application No. 70080/85, WO 94/00977), the electroporation method (Japanese Published Unexamined Patent Application No. 251887/85), the particle gun method (Japanese Patents 2606856 and 2517813), and the like.
- [0228] The transformant of the present invention includes a transformant containing the polypeptide of the present invention per se rather than as a recombinant vector, that is, a transformant containing the polypeptide of the present invention which is integrated into a chromosome of the host, in addition to the transformant containing the above recombinant vector.
  - [0229] When expressed in yeasts, animal cells, insect cells or plant cells, a glycopolypeptide or glycosylated polypeptide can be obtained.
  - [0230] The polypeptide can be produced by culturing the thus obtained transformant of the present invention in a culture medium to produce and accumulate the polypeptide of the present invention or any polypeptide expressed under the control of an EMF of the present invention, and recovering the polypeptide from the culture.
  - [0231] Culturing of the transformant of the present invention in a culture medium is carried out according to the conventional method as used in culturing of the host.
  - [0232] When the transformant of the present invention is obtained using a prokaryote, such as Escherichia coli or the like, or a eukaryote, such as yeast or the like, as the host, the transformant is cultured.
  - [0233] Any of a natural medium and a synthetic medium can be used, so long as it contains a carbon source, a nitrogen source, an inorganic salt and the like which can be assimilated by the transformant and can perform culturing of the transformant efficiently.

40

50

55

- [0234] Examples of the carbon source include those which can be assimilated by the transformant, such as carbohydrates (for example, glucose, fructose, sucrose, molasses containing them, starch, starch hydrolysate, and the like), organic acids (for example, acetic acid, propionic acid, and the like), and alcohols (for example, ethanol, propanol, and the like).
- [0235] Examples of the nitrogen source include ammonia, various ammonium salts of inorganic acids or organic acids (for example, ammonium chloride, ammonium sulfate, ammonium acetate, ammonium phosphate, and the like), other nitrogen-containing compounds, peptone, meat extract, yeast extract, corn steep liquor, casein hydrolysate, soybean meal and soybean meal hydrolysate, various fermented cells and hydrolysates thereof, and the like.
  - [0236] Examples of inorganic salt include potassium dihydrogen phosphate, dipotassium hydrogen phosphate, magnesium phosphate, magnesium sulfate, sodium chloride, ferrous sulfate, manganese sulfate, copper sulfate, calcium carbonate, and the like.
  - [0237] The culturing is carried out under aerobic conditions by shaking culture, submerged-aeration stirring culture or the like. The culturing temperature is preferably from 15 to 40°C, and the culturing time is generally from 16 hours to 7 days. The pH of the medium is preferably maintained at 3.0 to 9.0 during the culturing. The pH can be adjusted using an inorganic or organic acid, an alkali solution, urea, calcium carbonate, ammonia, or the like.
  - [0238] Also, antibiotics, such as ampicillin, tetracycline, and the like, can be added to the medium during the culturing, if necessary.
  - [0239] When a microorganism transformed with a recombinant vector containing an inducible promoter is cultured,

an inducer can be added to the medium, if necessary.

5

10

20

25

30

35

50

[0240] For example, isopropyl-β-D-thiogalactopyranoside (IPTG) or the like can be added to the medium when a microorganism transformed with a recombinant vector containing *lac* promoter is cultured, or indoleacrylic acid (IAA) or the like can by added thereto when a microorganism transformed with an expression vector containing *trp* promoter is cultured.

[0241] Examples of the medium used in culturing a transformant obtained using animal cells as the host cells include RPMI 1640 medium (*The Journal of the American Medical Association, 199.* 519 (1967)), Eagle's MEM medium (*Science, 122*: 501 (1952)), Dulbecco's modified MEM medium (*Virology, 8,* 396 (1959)), 199 Medium (*Proceeding of the Society for the Biological Medicine, 73*:1 (1950)), the above-described media to which fetal calf serum has been added, and the like.

[0242] The culturing is carried out generally at a pH of 6 to 8 and a temperature of 30 to 40°C in the presence of 5% CO<sub>2</sub> for 1 to 7 days.

[0243] Also, if necessary, antibiotics, such as kanamycin, penicillin, and the like, can be added to the medium during the culturing.

[0244] Examples of the medium used in culturing a transformant obtained using insect cells as the host cells include TNM-FH medium (manufactured by Pharmingen), Sf-900 II SFM (manufactured by Life Technologies), ExCell 400 and ExCell 405 (manufactured by JRH Biosciences), Grace's Insect Medium (Nature, 195: 788 (1962)), and the like.

[0245] The culturing is carried out generally at a pH of 6 to 7 and a temperature of 25 to 30°C for 1 to 5 days.

[0246] Additionally, antibiotics, such as gentamicin and the like, can be added to the medium during the culturing, if necessary.

[0247] A transformant obtained by using a plant cell as the host cell can be used as the cell or after differentiating to a plant cell or organ. Examples of the medium used in the culturing of the transformant include Murashige and Skoog (MS) medium, White medium, media to which a plant hormone, such as auxin, cytokinine, or the like has been added, and the like.

[0248] The culturing is carried out generally at a pH of 5 to 9 and a temperature of 20 to 40°C for 3 to 60 days.

[0249] Also, antibiotics, such as kanamycin, hygromycin and the like, can be added to the medium during the culturing, if necessary.

[0250] As described above, the polypeptide can be produced by culturing a transformant derived from a microorganism, animal cell or plant cell containing a recombinant vector to which a DNA encoding the polypeptide of the present invention has been inserted according to the general culturing method to produce and accumulate the polypeptide, and recovering the polypeptide from the culture.

[0251] The process of gene expression may include secretion of the encoded protein production or fusion protein expression and the like in accordance with the methods described in *Molecular Cloning*, 2nd ed., in addition to direct expression.

[0252] The method for producing the polypeptide of the present invention includes a method of intracellular expression in a host cell, a method of extracellular secretion from a host cell, or a method of production on a host cell membrane outer envelope. The method can be selected by changing the host cell employed or the structure of the polypeptide produced.

[0253] When the polypeptide of the present invention is produced in a host cell or on a host cell membrane outer envelope, the polypeptide can be positively secreted extracellularly according to, for example, the method of Paulson et al. (J. Biol. Chem., 264: 17619 (1989)), the method of Lowe et al. (Proc. Natl. Acad. Sci. USA, 86: 8227 (1989); Genes Develop., 4: 1288 (1990)), and/or the methods described in Japanese Published Unexamined Patent Application No. 336963/93, WO 94/23021, and the like.

[0254] Specifically, the polypeptide of the present invention can be positively secreted extracellularly by expressing it in the form that a signal peptide has been added to the foreground of a polypeptide containing an active site of the polypeptide of the present invention according to the recombinant DNA technique.

[0255] Furthermore, the amount produced can be increased using a gene amplification system, such as by use of a dihydrofolate reductase gene or the like according to the method described in Japanese Published Unexamined Patent Application No. 227075/90.

[0256] Moreover, the polypeptide of the present invention can be produced by a transgenic animal individual (transgenic nonhuman animal) or plant individual (transgenic plant).

[0257] When the transformant is the animal individual or plant individual, the polypeptide of the present invention can be produced by breeding or cultivating it so as to produce and accumulate the polypeptide, and recovering the polypeptide from the animal individual or plant individual.

[0258] Examples of the method for producing the polypeptide of the present invention using the animal individual include a method for producing the polypeptide of the present invention in an animal developed by inserting a gene according to methods known to those of ordinary skill in the art (American Journal of Clinical Nutrition, 63: 639S (1996), American Journal of Clinical Nutrition, 63: 627S (1996), Bio/Technology, 9: 830 (1991)).

[0259] In the animal individual, the polypeptide can be produced by breeding a transgenic nonhuman animal to which the DNA encoding the polypeptide of the present invention has been inserted to produce and accumulate the polypeptide in the animal, and recovering the polypeptide from the animal. Examples of the production and accumulation place in the animal include milk (Japanese Published Unexamined Patent Application No. 309192/88), egg and the like of the animal. Any promoter can be used, so long as it can be expressed in the animal. Suitable examples include an  $\alpha$ -casein promoter, a  $\beta$ -casein promoter, a  $\beta$ -lactoglobulin promoter, a whey acidic protein promoter, and the like, which are specific for mammary glandular cells.

[0260] Examples of the method for producing the polypeptide of the present invention using the plant individual include a method for producing the polypeptide of the present invention by cultivating a transgenic plant to which the DNA encoding the protein of the present invention by a known method (*Tissue Culture, 20* (1994), *Tissue Culture, 21* (1994), *Trends in Biotechnology, 15:* 45 (1997)) to produce and accumulate the polypeptide in the plant, and recovering the polypeptide from the plant.

[0261] The polypeptide according to the present invention can also be obtained by translation in vitro.

5

10

15

20

25

30

35

40

45

50

55

[0262] The polypeptide of the present invention can be produced by a translation system *in vitro*. There are, for example, two *in vitro* translation methods which may be used, namely, a method using RNA as a template and another method using DNA as a template. The template RNA includes the whole RNA, mRNA, an *in vitro* transcription product, and the like. The template DNA includes a plasmid containing a transcriptional promoter and a target gene integrated therein and downstream of the initiation site, a PCR/RT-PCR product and the like. To select the most suitable system for the *in vitro* translation, the origin of the gene encoding the protein to be synthesized (prokaryotic cell/eucaryotic cell), the type of the template (DNA/RNA), the purpose of using the synthesized protein and the like should be considered. *In vitro* translation kits having various characteristics are commercially available from many companies (Boehringer Mannheim, Promega, Stratagene, or the like), and every kit can be used in producing the polypeptide according to the present invention.

[0263] Transcription/translation of a DNA nucleotide sequence cloned into a plasmid containing a T7 promoter can be carried out using an *in vitro* transcription/translation system *E. coli* T7 S30 Extract System for Circular DNA (manufactured by Promega, catalogue No. L1130). Also, transcription/translation using, as a template, a linear prokaryotic DNA of a supercoil non-sensitive promoter, such as *lac*UV5, *tac*, λPL(con), λPL, or the like, can be carried out using an *in vitro* transcription/translation system *E. coli* S30 Extract System for Linear Templates (manufactured by Promega, catalogue No. L1030). Examples of the linear prokaryotic DNA used as a template include a DNA fragment, a PCR-amplified DNA product, a duplicated oligonucleotide ligation, an *in vitro* transcriptional RNA, a prokaryotic RNA, and the like.

[0264] In addition to the production of the polypeptide according to the present invention, synthesis of a radioactive labeled protein, confirmation of the expression capability of a cloned gene, analysis of the function of transcriptional reaction or translation reaction, and the like can be carried out using this system.

[0265] The polypeptide produced by the transformant of the present invention can be isolated and purified using the general method for isolating and purifying an enzyme. For example, when the polypeptide of the present invention is expressed as a soluble product in the host cells, the cells are collected by centrifugation after cultivation, suspended in an aqueous buffer, and disrupted using an ultrasonicator, a French press, a Manton Gaulin homogenizer, a Dynomill, or the like to obtain a cell-free extract. From the supernatant obtained by centrifuging the cell-free extract, a purified product can be obtained by the general method used for isolating and purifying an enzyme, for example, solvent extraction, salting out using ammonium sulfate or the like, desalting, precipitation using an organic solvent, anion exchange chromatography using a resin, such as S-Sepharose FF (manufactured by Mitsubishi Chemical) or the like, cation exchange chromatography using a resin, such as S-Sepharose FF (manufactured by Pharmacia) or the like, hydrophobic chromatography using a resin, such as butyl sepharose, phenyl sepharose or the like, gel filtration using a molecular sieve, affinity chromatography, chromatofocusing, or electrophoresis, such as isoelectronic focusing or the like, alone or in combination thereof.

[0266] When the polypeptide is expressed as an insoluble product in the host cells, the cells are collected in the same manner, disrupted and centrifuged to recover the insoluble product of the polypeptide as the precipitate fraction. Next, the insoluble product of the polypeptide is solubilized with a protein denaturing agent. The solubilized solution is diluted or dialyzed to lower the concentration of the protein denaturing agent in the solution. Thus, the normal configuration of the polypeptide is reconstituted. After the procedure, a purified product of the polypeptide can be obtained by a purification/isolation method similar to the above.

[0267] When the polypeptide of the present invention or its derivative (for example, a polypeptide formed by adding a sugar chain thereto) is secreted out of cells, the polypeptide or its derivative can be collected in the culture supernatant. Namely, the culture supernatant is obtained by treating the culture medium in a treatment similar to the above (for example, centrifugation). Then, a purified product can be obtained from the culture medium using a purification/isolation method similar to the above.

[0268] The polypeptide obtained by the above method is within the scope of the polypeptide of the present invention,

and examples include a polypeptide encoded by a polynucleotide comprising the nucleotide sequence selected from SEQ ID NOS:2 to 3431, and a polypeptide comprising an amino acid sequence represented by any one of SEQ ID NOS:3502 to 6931.

[0269] Furthermore, a polypeptide comprising an amino acid sequence in which at least one amino acids is deleted, replaced, inserted or added in the amino acid sequence of the polypeptide and having substantially the same activity as that of the polypeptide is included in the scope of the present invention. The term "substantially the same activity as that of the polypeptide" means the same activity represented by the inherent function, enzyme activity or the like possessed by the polypeptide which has not been deleted, replaced, inserted or added. The polypeptide can be obtained using a method for introducing part-specific mutation(s) described in, for example, *Molecular Cloning*, 2nd ed., *Current Protocols in Molecular Biology, Nuc. Acids. Res.*, 10: 6487 (1982), *Proc. Natl. Acad. Sci. USA*, 79: 6409 (1982), *Gene, 34*: 315 (1985), *Nuc. Acids. Res.*, 13: 4431 (1985), *Proc. Natl. Acad. Sci. USA*, 82: 488 (1985) and the like. For example, the polypeptide can be obtained by introducing mutation(s) to DNA encoding a polypeptide having the amino acid sequence represented by any one of SEQ ID NOS:3502 to 6931. The number of the amino acids which are deleted, replaced, inserted or added is not particularly limited; however, it is usually 1 to the order of tens, preferably 1 to 20, more preferably 1 to 10, and most preferably 1 to 5, amino acids.

[0270] The at least one amino acid deletion, replacement, insertion or addition in the amino acid sequence of the polypeptide of the present invention is used herein to refer to that at least one amino acid is deleted, replaced, inserted or added to at one or plural positions in the amino acid sequence. The deletion, replacement, insertion or addition may be caused in the same amino acid sequence simultaneously. Also, the amino acid residue replaced, inserted or added can be natural or non-natural. Examples of the natural amino acid residue include L-alanine, L-asparagine, L-asparatic acid, L-glutamine, L-glutamic acid, glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine, L-valine, L-cysteine, and the like.

[0271] Herein, examples of amino acid residues which are replaced with each other are shown below. The amino acid residues in the same group can be replaced with each other.

Group A:

5

10

15

20

25

30

40

50

55

[0272] leucine, isoleucine, norleucine, valine, norvaline, alanine, 2-aminobutanoic acid, methionine, O-methylserine, t-butylglycine, t-butylalanine, cyclohexylalanine;

Group B:

[0273] asparatic acid, glutamic acid, isoasparatic acid, isoglutamic acid, 2-aminoadipic acid, 2-aminosuberic acid;

35 Group C:

[0274] asparagine, glutamine;

Group D:

[0275] lysine, arginine, ornithine, 2,4-diaminobutanoic acid, 2,3-diaminopropionic acid;

Group E:

45 [0276] proline, 3-hydroxyproline, 4-hydroxyproline;

Group F:

[0277] serine, threonine, homoserine;

Group G:

[0278] phenylalanine, tyrosine.

[0279] Also, in order that the resulting mutant polypeptide has substantially the same activity as that of the polypeptide which has not been mutated, it is preferred that the mutant polypeptide has a homology of 60% or more, preferably 80% or more, and particularly preferably 95% or more, with the polypeptide which has not been mutated, when calculated, for example, using default (initial setting) parameters by a homology searching software, such as BLAST, FASTA, or the like.

[0280] Also, the polypeptide of the present invention can be produced by a chemical synthesis method, such as Fmoc (fluorenylmethyloxycarbonyl) method, tBoc (t-butyloxycarbonyl) method, or the like. It can also be synthesized using a peptide synthesizer manufactured by Advanced ChemTech, Perkin-Elmer, Pharmacia, Protein Technology Instrument, Synthecell-Vega, PerSeptive, Shimadzu Corporation, or the like.

[0281] The transformant of the present invention can be used for objects other than the production of the polypeptide of the present invention.

[0282] Specifically, at least one component selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogues thereof can be produced by culturing the transformant containing the polynucleotide or recombinant vector of the present invention in a medium to produce and accumulate at least one component selected from amino acids, nucleic acids, vitamins, saccharides, organic acids, and analogues thereof, and recovering the same from the medium.

[0283] The biosynthesis pathways, decomposition pathways and regulatory mechanisms of physiologically active substances such as amino acids, nucleic acids, vitamins, saccharides, organic acids and analogues thereof differ from organism to organism. The productivity of such a physiologically active substance can be improved using these differences, specifically by introducing a heterogeneous gene relating to the biosynthesis thereof. For example, the content of lysine, which is one of the essential amino acids, in a plant seed was improved by introducing a synthase gene derived from a bacterium (WO 93/19190). Also, arginine is excessively produced in a culture by introducing an arginine synthase gene derived from *Escherichia coli* (Japanese Examined Patent Publication 23750/93).

[0284] To produce such a physiologically active substance, the transformant according to the present invention can be cultured by the same method as employed in culturing the transformant for producing the polypeptide of the present invention as described above. Also, the physiologically active substance can be recovered from the culture medium in combination with, for example, the ion exchange resin method, the precipitation method and other known methods. [0285] Examples of methods known to one of ordinary skill in the art include electroporation, calcium transfection, the protoplast method, the method using a phage, and the like, when the host is a bacterium; and microinjection, calcium phosphate transfection, the positively charged lipid-mediated method and the method using a virus, and the like, when the host is a eukaryote (*Molecular Cloning*, 2nd ed.; Spector *et al.*, *Cells/a laboratory manual*, Cold Spring Harbour Laboratory Press, 1998)). Examples of the host include prokaryotes, lower eukaryotes (for example, yeasts), higher eukaryotes (for example, mammals), and cells isolated therefrom. As the state of a recombinant polynucleotide fragment present in the host cells, it can be integrated into the chromosome of the host. Alternatively, it can be integrated into a factor (for example, a plasmid) having an independent replication unit outside the chromosome. These transformants are usable in producing the polypeptides of the present invention encoded by the ORF of the genome of *Corynebacterium glutamicum*, the polynucleotides of the present invention and fragments thereof. Alternatively, they can be used in producing arbitrary polypeptides under the regulation by an EMF of the present invention.

11. Preparation of antibody recognizing the polypeptide of the present invention

[0286] An antibody which recognizes the polypeptide of the present invention, such as a polyclonal antibody, a monoclonal antibody, or the like, can be produced using, as an antigen, a purified product of the polypeptide of the present invention or a partial fragment polypeptide of the polypeptide or a peptide having a partial amino acid sequence of the polypeptide of the present invention.

(1) Production of polyclonal antibody

10

15

20

25

30

35

40

50

55

[0287] A polyclonal antibody can be produced using, as an antigen, a purified product of the polypeptide of the present invention, a partial fragment polypeptide of the polypeptide, or a peptide having a partial amino acid sequence of the polypeptide of the present invention, and immunizing an animal with the same.

[0288] Examples of the animal to be immunized include rabbits, goats, rats, mice, hamsters, chickens and the like.

[0289] A dosage of the antigen is preferably 50 to 100 µg per animal.

[0290] When the peptide is used as the antigen, it is preferably a peptide covalently bonded to a carrier protein, such as keyhole limpet haemocyanin, bovine thyroglobulin, or the like. The peptide used as the antigen can be synthesized by a peptide synthesizer.

[0291] The administration of the antigen is, for example, carried out 3 to 10 times at the intervals of 1 or 2 weeks after the first administration. On the 3rd to 7th day after each administration, a blood sample is collected from the venous plexus of the eyeground, and it is confirmed that the serum reacts with the antigen by the enzyme immunoassay (Enzyme-linked Immunosorbent Assay (ELISA), Igaku Shoin (1976); Antibodies - A Laboratory Manual, Cold Spring Harbor Laboratory (1988)) or the like.

[0292] Serum is obtained from the immunized non-human mammal with a sufficient antibody titer against the antigen used for the immunization, and the serum is isolated and purified to obtain a polyclonal antibody.

[0293] Examples of the method for the isolation and purification include centrifugation, salting out by 40-50% saturated ammonium sulfate, caprylic acid precipitation (*Antibodies, A Laboratory manual,* Cold Spring Harbor Laboratory (1988)), or chromatography using a DEAE-Sepharose column, an anion exchange column, a protein A- or G-column, a gel filtration column, and the like, alone or in combination thereof, by methods known to those of ordinary skill in the art.

(2) Production of monoclonal antibody

5

25

30

35

40

45

50

**55** 

- (a) Preparation of antibody-producing cell
- [0294] A rat having a serum showing an enough antibody titer against a partial fragment polypeptide of the polypeptide of the present invention used for immunization is used as a supply source of an antibody-producing cell.
  - [0295] On the 3rd to 7th day after the antigen substance is finally administered the rat showing the antibody titer, the spleen is excised.
  - [0296] The spleen is cut to pieces in MEM medium (manufactured by Nissui Pharmaceutical), loosened using a pair of forceps, followed by centrifugation at 1,200 rpm for 5 minutes, and the resulting supernatant is discarded.
  - [0297] The spleen in the precipitated fraction is treated with a Tris-ammonium chloride buffer (pH 7.65) for 1 to 2 minutes to eliminate erythrocytes and washed three times with MEM medium, and the resulting spleen cells are used as antibody-producing cells.
- 20 (b) Preparation of myeloma cells

[0298] As myeloma cells, an established cell line obtained from mouse or rat is used. Examples of useful cell lines include those derived from a mouse, such as P3-X63Ag8-U1 (hereinafter referred to as "P3-U1") (*Curr. Topics in Microbiol. Immunol., 81*: 1 (1978); *Europ. J. Immunol., 6*: 511 (1976)); SP2/O-Agl4 (SP-2) (*Nature, 276*: 269 (1978)): P3-X63-Ag8653 (653) (*J. Immunol., 123*: 1548 (1979)); P3-X63-Ag8 (X63) cell line (*Nature, 256*: 495 (1975)), and the like, which are 8-azaguanine-resistant mouse (BALB/c) myeloma cell lines. These cell lines are subcultured in 8-azaguanine medium (medium in which, to a medium obtained by adding 1.5 mmol/l glutamine,  $5 \times 10^{-5}$  mol/l 2-mercaptoethanol, 10 µg/ml gentamicin and 10% fetal calf serum (FCS) (manufactured by CSL) to RPMI-1640 medium (hereinafter referred to as the "normal medium"), 8-azaguanine is further added at 15 µg/ml) and cultured in the normal medium 3 or 4 days before cell fusion, and  $2 \times 10^7$  or more of the cells are used for the fusion.

(c) Production of hybridoma

[0299] The antibody-producing cells obtained in (a) and the myeloma cells obtained in (b) are washed with MEM medium or PBS (disodium hydrogen phosphate: 1.83 g, sodium dihydrogen phosphate: 0.21 g, sodium chloride: 7.65 g, distilled water: 1 liter, pH: 7.2) and mixed to give a ratio of antibody-producing cells: myeloma cells = 5:1 to 10: 1. followed by centrifugation at 1,200 rpm for 5 minutes, and the supernatant is discarded.

[0300] The cells in the resulting precipitated fraction were thoroughly loosened, 0.2 to 1 ml of a mixed solution of 2 g of polyethylene glycol-1000 (PEG-1000), 2 ml of MEM medium and 0.7 ml of dimethylsulfoxide (DMSO) per 10<sup>8</sup> antibody-producing cells is added to the cells under stirring at 37°C, and then 1 to 2 ml of MEM medium is further added thereto several times at 1 to 2 minute intervals.

[0301] After the addition, MEM medium is added to give a total amount of 50 ml. The resulting prepared solution is centrifuged at 900 rpm for 5 minutes, and then the supernatant is discarded. The cells in the resulting precipitated fraction were gently loosened and then gently suspended in 100 ml of HAT medium (the normal medium to which 10<sup>-4</sup> mol/l hypoxanthine, 1.5×10<sup>-5</sup> mol/l thymidine and 4×10<sup>-7</sup> mol/l aminopterin have been added) by repeated drawing up into and discharging from a measuring pipette.

[0302] The suspension is poured into a 96 well culture plate at 100  $\mu$ l/well and cultured at 37°C for 7 to 14 days in a 5% CO<sub>2</sub> incubator.

[0303] After culturing, a part of the culture supernatant is recovered, and a hybridoma which specifically reacts with a partial fragment polypeptide of the polypeptide of the present invention is selected according to the enzyme immunoassay described in *Antibodies, A Laboratory manual*, Cold Spring Harbor Laboratory, Chapter 14 (1998) and the like. [0304] A specific example of the enzyme immunoassay is described below.

[0305] The partial fragment polypeptide of the polypeptide of the present invention used as the antigen in the immunization is spread on a suitable plate, is allowed to react with a hybridoma culturing supernatant or a purified antibody obtained in (d) described below as a first antibody, and is further allowed to react with an anti-rat or anti-mouse immunoglobulin antibody labeled with an enzyme, a chemical luminous substance, a radioactive substance or the like as a second antibody for reaction suitable for the labeled substance. A hybridoma which specifically reacts with the polypeptide of the present invention is selected as a hybridoma capable of producing a monoclonal antibody of the present

invention.

5

10

15

20

25

35

40

[0306] Cloning is repeated using the hybridoma twice by limiting dilution analysis (HT medium (a medium in which aminopterin has been removed from HAT medium) is firstly used, and the normal medium is secondly used), and a hybridoma which is stable and contains a sufficient amount of antibody titer is selected as a hybridoma capable of producing a monoclonal antibody of the present invention.

### (d) Preparation of monoclonal antibody

[0307] The monoclonal antibody-producing hybridoma cells obtained in (c) are injected intraperitoneally into 8- to 10-week-old mice or nude mice treated with pristane (intraperitoneal administration of 0.5 ml of 2,6,10,14-tetrameth-ylpentadecane (pristane), followed by 2 weeks of feeding) at  $5\times10^6$  to  $20\times10^6$  cells/animal. The hybridoma causes ascites tumor in 10 to 21 days.

[0308] The ascitic fluid is collected from the mice or nude mice, and centrifuged to remove solid contents at 3000 rpm for 5 minutes.

[0309] A monoclonal antibody can be purified and isolated from the resulting supernatant according to the method similar to that used in the polyclonal antibody.

[0310] The subclass of the antibody can be determined using a mouse monoclonal antibody typing kit or a rat monoclonal antibody typing kit. The polypeptide amount can be determined by the Lowry method or by calculation based on the absorbance at 280 nm.

[0311] The antibody obtained in the above is within the scope of the antibody of the present invention.

[0312] The antibody can be used for the general assay using an antibody, such as a radioactive material labeled immunoassay (RIA), competitive binding assay, an immunotissue chemical staining method (ABC method, CSA method, etc.), immunoprecipitation, Western blotting, ELISA assay, and the like (*An introduction to Radioimmunoassay and Related Techniques*, Elsevier Science (1986); *Techniques in Immunocytochemistry*, Academic Press, Vol. 1 (1982),

Vol. 2 (1983) & Vol. 3 (1985); Practice and Theory of Enzyme Immunoassays, Elsevier Science (1985); Enzyme-linked Immunosorbent Assay (ELISA), Igaku Shoin (1976); Antibodies - A Laboratory Manual, Cold Spring Harbor laboratory (1988); Monoclonal Antibody Experiment Manual, Kodansha Scientific (1987); Second Series Biochemical Experiment Course, Vol. 5, Immunobiochemistry Research Method, Tokyo Kagaku Dojin (1986)).

[0313] The antibody of the present invention can be used as it is or after being labeled with a label.

[0314] Examples of the label include radioisotope, an affinity label (e.g., biotin, avidin, or the like), an enzyme label (e.g., horseradish peroxidase, alkaline phosphatase, or the like), a fluorescence label (e.g., FITC, rhodamine, or the like), a label using a rhodamine atom, (*J. Histochem. Cytochem.*, 18: 315 (1970); Meth. Enzym., 62: 308 (1979); Immunol., 109: 129 (1972); J. Immunol., Meth., 13: 215 (1979)), and the like.

[0315] Expression of the polypeptide of the present invention, fluctuation of the expression, the presence or absence of structural change of the polypeptide, and the presence or absence in an organism other than coryneform bacteria of a polypeptide corresponding to the polypeptide can be analyzed using the antibody or the labeled antibody by the above assay, or a polypeptide array or proteome analysis described below.

[0316] Furthermore, the polypeptide recognized by the antibody can be purified by immunoaffinity chromatography using the antibody of the present invention.

# 12. Production and use of polypeptide array

### (1) Production of polypeptide array

[0317] A polypeptide array can be produced using the polypeptide of the present invention obtained in the above item 10 or the antibody of the present invention obtained in the above item 11.

[0318] The polypeptide array of the present invention includes protein chips, and comprises a solid support and the polypeptide or antibody of the present invention adhered to the surface of the solid support.

[0319] Examples of the solid support include plastic such as polycarbonate or the like; an acrylic resin, such as polyacrylamide or the like; complex carbohydrates, such as agarose, sepharose, or the like; silica; a silica-based material, carbon, a metal, inorganic glass, latex beads, and the like.

[0320] The polypeptides or antibodies according to the present invention can be adhered to the surface of the solid support according to the method described in *Biotechniques*, 27: 1258-61 (1999); *Molecular Medicine Today*, 5: 326-7 (1999); *Handbook of Experimental Immunology*, 4th edition, Blackwell Scientific Publications, Chapter 10 (1986); *Meth. Enzym.*, 34 (1974); *Advances in Experimental Medicine and Biology*, 42 (1974); U.S. Patent 4,681,870; U.S. Patent 4,282,287; U.S. Patent 4,762,881, or the like.

[0321] The analysis described herein can be efficiently performed by adhering the polypeptide or antibody of the present invention to the solid support at a high density, though a high fixation density is not always necessary.

#### (2) Use of polypeptide array

5

10

15

20

25

30

35

40

45

50

[0322] A polypeptide or a compound capable of binding to and interacting with the polypeptides of the present invention adhered to the array can be identified using the polypeptide array to which the polypeptides of the present invention have been adhered thereto as described in the above (1).

[0323] Specifically, a polypeptide or a compound capable of binding to and interacting with the polypeptides of the present invention can be identified by subjecting the polypeptides of the present invention to the following steps (i) to (iv):

- (i) preparing a polypeptide array having the polypeptide of the present invention adhered thereto by the method of the above (1);
- (ii) incubating the polypeptide immobilized on the polypeptide array together with at least one of a second polypeptide or compound;
- (iii) detecting any complex formed between the at least one of a second polypeptide or compound and the polypeptide immobilized on the array using, for example, a label bound to the at least one of a second polypeptide or compound, or a secondary label which specifically binds to the complex or to a component of the complex after unbound material has been removed; and
- (iv) analyzing the detection data.

[0324] Specific examples of the polypeptide array to which the polypeptide of the present invention has been adhered include a polypeptide array containing a solid support to which at least one of a polypeptide containing an amino acid sequence selected from SEQ ID NOS:3502 to 7001, a polypeptide containing an amino acid sequence in which at least one amino acids is deleted, replaced, inserted or added in the amino acid sequence of the polypeptide and having substantially the same activity as that of the polypeptide, a polypeptide containing an amino acid sequence having a homology of 60% or more with the amino acid sequences of the polypeptide and having substantially the same activity as that of the polypeptides, a partial fragment polypeptide, and a peptide comprising an amino acid sequence of a part of a polypeptide.

[0325] The amount of production of a polypeptide derived from coryneform bacteria can be analyzed using a polypeptide array to which the antibody of the present invention has been adhered in the above (1).

[0326] Specifically, the expression amount of a gene derived from a mutant of coryneform bacteria can be analyzed by subjecting the gene to the following steps (i) to (iv):

- (i) preparing a polypeptide array by the method of the above (1);
- (ii) incubating the polypeptide array (the first antibody) together with a polypeptide derived from a mutant of coryneform bacteria;
- (iii) detecting the polypeptide bound to the polypeptide immobilized on the array using a labeled second antibody of the present invention; and
- (iv) analyzing the detection data.

[0327] Specific examples of the polypeptide array to which the antibody of the present invention is adhered include a polypeptide array comprising a solid support to which at least one of an antibody which recognizes a polypeptide comprising an amino acid sequence selected from SEQ ID NOS:3502 to 7001, a polypeptide comprising an amino acid sequence in which at least one amino acids is deleted, replaced, inserted or added in the amino acid sequence of the polypeptide and having substantially the same activity as that of the polypeptide, a polypeptide comprising an amino acid sequence having a homology of 60% or more with the amino acid sequences of the polypeptide and having substantially the same activity as that of the polypeptides, a partial fragment polypeptide, or a peptide comprising an amino acid sequence of a part of a polypeptide.

[0328] A fluctuation in an expression amount of a specific polypeptide can be monitored using a polypeptide obtained in the time course of culture as the polypeptide derived from coryneform bacteria. The culturing conditions can be optimized by analyzing the fluctuation.

[0329] When a polypeptide derived from a mutant of coryneform bacteria is used, a mutated polypeptide can be detected.

- 13. Identification of useful mutation in mutant by proteome analysis
- [0330] Usually, the proteome is used herein to refer to a method wherein a polypeptide is separated by twodimensional electrophoresis and the separated polypeptide is digested with an enzyme, followed by identification of the polypeptide using a mass spectrometer (MS) and searching a data base.
  - [0331] The two dimensional electrophoresis means an electrophoretic method which is performed by combining two

electrophoretic procedures having different principles. For example, polypeptides are separated depending on molecular weight in the primary electrophoresis. Next, the gel is rotated by 90° or 180° and the secondary electrophoresis is carried out depending on isoelectric point. Thus, various separation patterns can be achieved (JIS K 3600 2474).

[0332] In searching the data base, the amino acid sequence information of the polypeptides of the present invention and the recording medium of the present invention provide for in the above items 2 and 8 can be used.

[0333] The proteome analysis of a coryneform bacterium and its mutant makes it possible to identify a polypeptide showing a fluctuation therebetween.

[0334] The proteome analysis of a wild type strain of coryneform bacteria and a production strain showing an improved productivity of a target product makes it possible to efficiently identify a mutation protein which is useful in breeding for improving the productivity of a target product or a protein of which expression amount is fluctuated.

[0335] Specifically, a wild type strain of coryneform bacteria and a lysine-producing strain thereof are each subjected to the proteome analysis. Then, a spot increased in the lysine-producing strain, compared with the wild type strain, is found and a data base is searched so that a polypeptide showing an increase in yield in accordance with an increase in the lysine productivity can be identified. For example, as a result of the proteome analysis on a wild type strain and a lysine-producing strain, the productivity of the catalase having the amino acid sequence represented by SEQ ID NO: 3785 is increased in the lysine-producing mutant.

[0336] As a result that a protein having a high expression level is identified by proteome analysis using the nucleotide sequence information and the amino acid sequence information, of the genome of the coryneform bacteria of the present invention, and a recording medium storing the sequences, the nucleotide sequence of the gene encoding this protein and the nucleotide sequence in the upstream thereof can be searched at the same time, and thus, a nucleotide sequence having a high expression promoter can be efficiently selected.

[0337] In the proteome analysis, a spot on the two-dimentional electrophoresis gel showing a fluctuation is sometimes derived from a modified protein. However, the modified protein can be efficiently identified using the recording medium storing the nucleotide sequence information, the amino acid sequence information, of the genome of coryneform bacteria, and the recording medium storing the sequences, according to the present invention.

[0338] Moreover, a useful mutation point in a useful mutant can be easily specified by searching a nucleotide sequence (nucleotide sequence of promoters, ORF, or the like) relating to the thus identified protein using a recording medium storing the nucleotide sequence information and the amino acid sequence information, of the genome of coryneform bacteria of the present invention, and a recording medium storing the sequences and using a primer designed on the basis of the detected nucleotide sequence. As a result that the useful mutation point is specified, an industrially useful mutant having the useful mutation or other useful mutation derived therefrom can be easily bred.

[0339] The present invention will be explained in detail below based on Examples. However, the present invention is not limited thereto.

# Example 1

5

10

15

20

25

30

35

40

45

50

55

Determination of the full nucleotide sequence of genome of Corynebacterium glutamicum

[0340] The full nucleotide sequence of the genome of *Corynebacterium glutamicum* was determined based on the whole genome shotgun method (*Science, 269*: 496-512 (1995)). In this method, a genome library was prepared and the terminal sequences were determined at random. Subsequently, these sequences were ligated on a computer to cover the full genome. Specifically, the following procedure was carried out.

(1) Preparation of genome DNA of Corynebacterium glutamicum ATCC 13032

[0341] Corynebacterium glutamicum ATCC 13032 was cultured in BY medium (7 g/l meat extract, 10 g/l peptone, 3 g/l sodium chloride, 5 g/l yeast extract, pH 7.2) containing 1% of glycine at 30°C overnight and the cells were collected by centrifugation. After washing with STE buffer (10.3% sucrose, 25 mmol/l Tris hydrochloride, 25 mmol/l EDTA, pH 8.0), the cells were suspended in 10 ml of STE buffer containing 10 mg/ml lysozyme, followed by gently shaking at 37°C for 1 hour. Then, 2 ml of 10% SDS was added thereto to lyse the cells, and the resultant mixture was maintained at 65°C for 10 minutes and then cooled to room temperature. Then, 10 ml of Tris-neutralized phenol was added thereto, followed by gently shaking at room temperature for 30 minutes and centrifugation (15,000 × g, 20 minutes, 20°C). The aqueous layer was separated and subjected to extraction with phenol/chloroform and extraction with chloroform (twice) in the same manner. To the aqueous layer, 3 mol/l sodium acetate solution (pH 5.2) and isopropanol were added at 1/10 times volume and twice volume, respectively, followed by gently stirring to precipitate the genome DNA. The genome DNA was dissolved again in 3 ml of TE buffer (10 mmol/l Tris hydrochloride, 1 mmol/l EDTA, pH 8.0) containing 0.02 mg/ml of RNase and maintained at 37°C for 45 minutes. The extractions with phenol, phenol/chloroform and chloroform were carried out successively in the same manner as the above. The genome DNA was subjected to iso-

propanol precipitation. The thus formed genome DNA precipitate was washed with 70% ethanol three times, followed by air-drying, and dissolved in 1.25 ml of TE buffer to give a genome DNA solution (concentration: 0.1 mg/ml).

(2) Construction of a shotgun library

5

10

15

20

25

30

50

55

[0342] TE buffer was added to 0.01 mg of the thus prepared genome DNA of *Corynebacterium glutamicum* ATCC 13032 to give a total volume of 0.4 ml, and the mixture was treated with a sonicator (Yamato Powersonic Model 150) at an output of 20 continuously for 5 seconds to obtain fragments of 1 to 10 kb. The genome fragments were bluntended using a DNA blunting kit (manufactured by Takara Shuzo) and then fractionated by 6% polyacrylamide gel electrophoresis. Genome fragments of 1 to 2 kb were cut out from the gel, and 0.3 ml MG elution buffer (0.5 mol/l ammonium acetate, 10 mmol/l magnesium acetate, 1 mmol/l EDTA, 0.1% SDS) was added thereto, followed by shaking at 37°C overnight to elute DNA. The DNA eluate was treated with phenol/chloroform, and then precipitated with ethanol to obtain a genome library insert. The total insert and 500 ng of pUC18 *Smal*/BAP (manufactured by Amersham Pharmacia Biotech) were ligated at 16°C for 40 hours.

[0343] The ligation product was precipitated with ethanol and dissolved in 0.01 ml of TE buffer. The ligation solution (0.001 ml) was introduced into 0.04 ml of *E. coli* ELECTRO MAX DH10B (manufactured by Life Technologies) by the electroporation under conditions according to the manufacture's instructions. The mixture was spread on LB plate medium (LB medium (10 g/l bactotrypton, 5 g/l yeast extract, 10 g/l sodium chloride, pH 7.0) containing 1.6% of agar) containing 0.1 mg/ml ampicillin, 0.1 mg/ml X-gal and 1 mmol/l isopropyl-β-D-thiogalactopyranoside (IPTG) and cultured at 37°C overnight.

[0344] The transformant obtained from colonies formed on the plate medium was stationarily cultured in a 96-well titer plate having 0.05 ml of LB medium containing 0.1 mg/ml ampicillin at 37°C overnight. Then, 0.05 ml of LB medium containing 20% glycerol was added thereto, followed by stirring to obtain a glycerol stock.

(3) Construction of cosmid library

[0345] About 0.1 mg of the genome DNA of *Corynebacterium glutamicum* ATCC 13032 was partially digested with *Sau*3AI (manufactured by Takara Shuzo) and then ultracentrifuged (26,000 rpm, 18 hours, 20°C) under 10 to 40% sucrose density gradient obtained using 10% and 40% sucrose buffers (1 mol/l NaCl, 20 mmol/l Tris hydrochloride, 5 mmol/l EDTA, 10% or 40% sucrose, pH 8.0). After the centrifugation, the solution thus separated was fractionated into tubes at 1 ml in each tube. After confirming the DNA fragment length of each fraction by agarose gel electrophoresis, a fraction containing a large amount of DNA fragment of about 40 kb was precipitated with ethanol.

[0346] The DNA fragment was ligated to the BamHI site of superCos1 (manufactured by Stratagene) in accordance with the manufacture's instructions. The ligation product was incorporated into Escherichia coli XL-1-BlueMR strain (manufactured by Stratagene) using Gigapack III Gold Packaging Extract (manufactured by Stratagene) in accordance with the manufacture's instructions. The Escherichia coli was spread on LB plate medium containing 0.1 mg/ml ampicillin and cultured therein at 37°C overnight to isolate colonies. The resulting colonies were stationarily cultured at 37°C overnight in a 96-well titer plate containing 0.05 ml of the LB medium containing 0.1 mg/ml ampicillin in each well. LB medium containing 20% glycerol (0.05 ml) was added thereto, followed by stirring to obtain a glycerol stock.

(4) Determination of nucleotide sequence

(4-1) Preparation of template

[0347] The full nucleotide sequence of *Corynebacterium glutamicum* ATCC 13032 was determined mainly based on the whole genome shotgun method. The template used in the whole genome shotgun method was prepared by the PCR method using the library prepared in the above (2).

[0348] Specifically, the clone derived from the whole genome shotgun library was inoculated using a replicator (manufactured by GENETIX) into each well of a 96-well plate containing the LB medium containing 0.1 mg/ml of ampicillin at 0.08 ml per each well and then stationarily cultured at 37°C overnight.

[0349] Next, the culturing solution was transported using a copy plate (manufactured by Tokken) into a 96-well reaction plate (manufactured by PE Biosystems) containing a PCR reaction solution (TaKaRa Ex Taq (manufactured by Takara Shuzo)) at 0.08 ml per each well. Then, PCR was carried out in accordance with the protocol by Makino *et al.* (*DNA Research, 5*: 1-9 (1998)) using GeneAmp PCR System 9700 (manufactured by PE Biosystems) to amplify the inserted fragment.

[0350] The excessive primers and nucleotides were eliminated using a kit for purifying a PCR production (manufactured by Amersham Pharmacia Biotech) and the residue was used as the template in the sequencing reaction.

[0351] Some nucleotide sequences were determined using a double-stranded DNA plasmid as a template.

- [0352] The double-stranded DNA plasmid as the template was obtained by the following method.
- [0353] The clone derived from the whole genome shotgun library was inoculated into a 24- or 96-well plate containing a 2× YT medium (16 g/l bactotrypton, 10 g/l yeast extract, 5 g/l sodium chloride, pH 7.0) containing 0.05 mg/ml ampicillin at 1.5 ml per each well and then cultured under shaking at 37°C overnight.
- 5 [0354] The double-stranded DNA plasmid was prepared from the culturing solution using an automatic plasmid preparing machine, KURABO PI-50 (manufactured by Kurabo Industries) or a multiscreen (manufactured by Millipore) in accordance with the protocol provided by the manufacturer.
  - [0355] To purify the double-stranded DNA plasmid using the multiscreen, Biomek 2000 (manufactured by Beckman Coulter) or the like was employed.
- 10 [0356] The thus obtained double-stranded DNA plasmid was dissolved in water to give a concentration of about 0.1 mg/ml and used as the template in sequencing.

### (4-2) Sequencing reaction

- 15 [0357] To 6 μl of a solution of ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (manufactured by PE Biosystems), an M13 regular direction primer (M13-21) or an M13 reverse direction primer (M13REV) (*DNA Research*, 5: 1-9 (1998) and the template prepared in the above (4-1) (the PCR product or the plasmid) were added to give 10 μl of a sequencing reaction solution. The primers and the templates were used in an amount of 1.6 pmol and an amount of 50 to 200 ng, respectively.
- 20 [0358] Dye terminator sequencing reaction of 45 cycles was carried out with GeneAmp PCR System 9700 (manufactured by PE Biosystems) using the reaction solution. The cycle parameter was determined in accordance with the manufacturer's instruction accompanying ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit. The sample was purified using MultiScreen HV plate (manufactured by Millipore) according to the manufacture's instructions. The thus purified reaction product was precipitated with ethanol, followed by drying, and then stored in the dark at -30°C.
  - [0359] The dry reaction product was analyzed by ABI PRISM 377 DNA Sequencer and ABI PRISM 3700 DNA Analyzer (both manufactured by PE Biosystems) each in accordance with the manufacture's instructions.
  - [0360] The data of about 50,000 sequences in total (i.e., about 42,000 sequences obtained using 377 DNA Sequencer and about 8,000 reactions obtained by 3700 DNA Analyser) were transferred to a server (Alpha Server 4100: manufactured by COMPAQ) and stored. The data of these about 50,000 sequences corresponded to 6 times as much as the genome size.

## (5) Assembly

30

50

- [0361] All operations were carried out on the basis of UNIX platform. The analytical data were output in Macintosh platform using X Window System. The base call was carried out using phred (The University of Washington). The vector sequence data was deleted using SPS Cross\_Match (manufactured by Southwest Parallel Software). The assembly was carried out using SPS phrap (manufactured by Southwest Parallel Software; a high-speed version of phrap (The University of Washington)). The contig obtained by the assembly was analyzed using a graphical editor, consed (The University of Washington). A series of the operations from the base call to the assembly were carried out simultaneously using a script phredPhrap attached to consed.
  - (6) Determination of nucleotide sequence in gap part
- [0362] Each cosmid in the cosmid library constructed in the above (3) was prepared by a method similar to the preparation of the double-stranded DNA plasmid described in the above (4-1). The nucleotide sequence at the end of the inserted fragment of the cosmid was determined by using ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (manufactured by PE Biosystems) according to the manufacture's instructions.
  - [0363] About 800 cosmid clones were sequenced at both ends to search a nucleotide sequence in the contig derived from the shotgun sequencing obtained in the above (5) coincident with the sequence. Thus, the linkage between respective cosmid clones and respective contigs were determined and mutual alignment was carried out. Furthermore, the results were compared with the physical map of *Corynebacterium glutamicum* ATCC 13032 (*Mol. Gen. Genet., 252*: 255-265 (1996) to carrying out mapping between the cosmids and the contigs.
  - [0364] The sequence in the region which was not covered with the contigs was determined by the following method.

    [0365] Clones containing sequences positioned at the ends of contigs were selected. Among these clones, about 1,000 clones wherein only one end of the inserted fragment had been determined were selected and the sequence at the opposite end of the inserted fragment was determined. A shotgun library clone or a cosmid clone containing the sequences at the respective ends of the inserted fragment in two contigs was identified, the full nucleotide sequence

of the inserted fragment of this clone was determined, and thus the nucleotide sequence of the gap part was determined. When no shotgun library clone or cosmid clone covering the gap part was available, primers complementary to the end sequences at the two contigs were prepared and the DNA fragment in the gap part was amplified by PCR. Then, sequencing was performed by the primer walking method using the amplified DNA fragment as a template or by the shotgun method in which the sequence of a shotgun clone prepared from the amplified DNA fragment was determined. Thus, the nucleotide sequence of the domain was determined.

[0366] In a region showing a low sequence precision, primers were synthesized using AUTOFINISH function and NAVIGATING function of consed (The University of Washington) and the sequence was determined by the primer walking method to improve the sequence precision. The thus determined full nucleotide sequence of the genome of Corynebacterium glutamicum ATCC 13032 strain is shown in SEQ ID NO:1.

### (7) Identification of ORF and presumption of its function

5

10

15

20

25

30

35

40

45

50

55

[0367] ORFs in the nucleotide sequence represented by SEQ ID NO:1 were identified according to the following method. First, the ORF regions were determined using software for identifying ORF, i.e., Glimmer, GeneMark and GeneMark.hmm on UNIX platform according to the respective manual attached to the software.

[0368] Based on the data thus obtained, ORFs in the nucleotide sequence represented by SEQ ID NO:1 were identified.

[0369] The putative function of an ORF was determined by searching the homology of the identified amino acid sequence of the ORF against an amino acid database consisting of protein-encoding domains derived from Swiss-Prot, PIR or Genpept database constituted by protein encoding domains derived from GenBank database, Frame Search (manufactured by Compugen), or by searching the homology of the identified amino acid sequence of the ORF against an amino acid database consisting of protein-encoding domains derived from Swiss-Prot, PIR or Genpept database constituted by protein encoding domains derived from GenBank database, BLAST. The nucleotide sequences of the thus determined ORFs are shown in SEQ ID NOS:2 to 3501, and the amino acid sequences encoded by these ORFs are shown in SEQ ID NOS:3502 to 7001.

[0370] In some cases of the sequence listings in the present invention, nucleotide sequences, such as TTG, TGT, GGT, and the like, other than ATG, are read as an initiating codon encoding Met.

[0371] Also, the preferred nucleotide sequences are SEQ ID NOS:2 to 355 and 357 to 3501, and the preferred amino acid sequences are shown in SEQ ID NOS:3502 to 3855 and 3857 to 7001

[0372] Table 1 shows the registration numbers in the above-described databases of sequences which were judged as having the highest homology with the nucleotide sequences of the ORFs as the results of the homology search in the amino acid sequences using the homology-searching software Frame Search (manufactured by Compugen), names of the genes of these sequences, the functions of the genes, and the matched length, identities and analogies compared with publicly known amino acid translation sequences. Moreover, the corresponding positions were confirmed via the alignment of the nucleotide sequence of an arbitrary ORF with the nucleotide sequence of SEQ ID NO: 1. Also, the positions of nucleotide sequences other than the ORFs (for example, ribosomal RNA genes, transfer RNA genes, IS sequences, and the like) on the genome were determined.

[0373] Fig. 1 shows the positions of typical genes of the Corynebacterium glutamicum ATCC 13032 on the genome.

	r	· · · · · · · · · · · · · · · · · · ·			- [	1						F				T		—Т					
5		Function	on protein DnaA		III beta chain	rotein (recF	nie	ise (ATP-					ressor			ınit A	Ibrane protein	in	ry protein, LysR		genesis protein	ein	
10		T.	replication initiation protein DnaA		DNA polymerase III beta chain	DNA replication protein (recF protein)	hypothetical protein	DNA topoisomerase (ATP- hydrolyzing)					NAGC/XYLR repressor			DNA gyrase subunit A	hypothetical membrane protein	hypothetical protein	bacterial regulatory protein, LysR type		cytochrome c biogenesis protein	hypothetical protein	repressor
15		Matched length (a.a.)	524		390	392	174	704					422	-		854	112	329	268		265	155	117
20		Similarity (%)	8.66		81.8	79.9	58.1	88.9					50.7			88.1	9.69	63.5	62.3		57.4	64.5	70.1
		Identity (%)	96.8		50.5	53.3	35.1	71.9					29.4			70.4	29.5	33.7	27.6		29.1	31.6	36.8
25	Table 1	us gene	avum dnaA		megmatis dnaN	megmatis recF	elicolor yreG	ıberculosis					uberculosis			uberculosis yrA	uberculosis	<12 yeiH	thermoluteolus		sulatus ccdA	com1	berculosis
30	Ta	Homologous gene	Brevibacterium flavum dnaA		Mycobacterium smegmatis dnaN	Mycobacterium smegmatis recF	Streptomyces coelicolor yreG	Mycobacterium tuberculosis H37Rv gyrB					Mycobacterium tuberculosis H37Rv			Mycobacterium tuberculosis H37Rv Rv0006 gyrA	Mycobacterium tuberculosis H37Rv Rv0007	Escherichia coli K12 yeiH	Hydrogenophilus thermoluteolus TH-1 cbbR		Rhodobacter capsulatus ccdA	Coxiella burnetii com1	Mycobacterium tuberculosis H37Rv Rv1846c
35 40		db Match	gsp:R98523		sp:DP3B_MYCSM	SP. RECF_MYCSM	sp.YREG_STRCO	pir.S44198					sp:YV11_MYCTU			sp.GYRA_MYCTU	pir.E70698	sp:YEIH_ECOLI	gp:A8042619_1		gp:AF156103_2	pir:A49232	pir.F70664
		ORF (bp)	1572	324	1182	1182 8	534	2133	996	699	510	441	1071	261	246	2568 8	342	1035	894	420	870	762	369
45		Terminal (nt)	1572	1597	3473	4766	5299	7486	8795	8628	10071	9474	10107	11263	11523	14398	14746	15209	17207	17670	17860	18736	20073
50		Initial (nt)	-	1920	2292	3585	4766	5354	7830	9466	9562	9914	11177	11523	11768	11831	14405	16243	16314	17251	18729	19497	19705
		SEQ NO. (a.a.)	3502	3503	3504	3505	3506	3507	3508	3509	3510	3511	3512	3513	3514	3515	3516	3517	3518	3519	3520	3521	3522
55		SEQ NO. (DNA)	2	3	4	5	9	7	80	6	5	11	12	13	14	15	16	17	18	19	20	21	22

EP 1 108 790 A2

	_	_																		
5		Function	hypothetical membrane protein	2,5-diketo-D-gluconic acid reductase	5'-nucleotidase precursor	5'-nucleotidase family protein	transposase	organic hydroperoxide detoxication enzyme	ATP-dependent DNA helicase		glucan 1,4-alpha-glucosidase	lipaprotein	ABC 3 transport family or integral membrane protein	iron(III) dicitrate transport ATP- biding protein	sugar ABC transporter, periplasmic sugar-binding protein	high affinity ribose transport protein	ribose transport ATP-binding protein	neurofilament subunit NF-180	peptidyl-prolyl cis-trans isomerase A	hypothetical membrane protein
15		Matched length (a.a.)	321	26	196	270	51	139	217		449	311	266	222	283	312	236	347	169	226
20		Similarity (%)	50.8	88.5	56.1	56.7	72.6	79.9	8.09		54.1	63.7	74.1	70.3	56.5	68.3	76.7	44.4	89.9	53.1
		identity (%)	24.9	65.4	27.0	27.0	52.9	51.8	32.7		26.7	28.9	34.6	39.2	25.8	30.5	32.2	23.6	79.9	29.2
25	inved)	ene	4	ATCC	us nutA	ans	atum ORF1	stris	ans recG		visiae	athiae	nes SF370	ecE	MSBB	rbsC	bsA		e H37RV	qgP
<i>30</i> <i>35</i>	Table 1 (continued)	Homologous gene	Mycobacterium leprae MLCB1788.18	Corynebacterium sp. ATCC 31090	Vibrio parahaemolyticus nutA	Deinococcus radiodurans DR0505	Corynebacterium striatum ORF1	Xanthomonas campestris phaseoli ohr	Thiobacillus ferrooxidans recG	-	Saccharomyces cerevisiae S288C YIR019C sta1	Erysipelothrix rhusiopathiae ewlA	Streptococcus pyogenes SF370 mtsC	Escherichia coli K12 fecE	Thermotoga maritima MSB8 TM0114	Escherichia coli K12 rbsC	Bacillus subtilis 168 rbsA	Petromyzon marinus	Mycobacterium leprae H37RV RV0009 ppiA	Bacillus subtilis 168 yqgP
40		db Match	gp:MLCB1788_6	pir:140838	sp:5NTD_VIBPA	gp:AE001909_7	prf.2513302C	prf.2413353A	sp:RECG_THIFE		sp:AMYH_YEAST	gp:ERU52850_1	gp:AF180520_3	sp:FECE_ECOLI	pir.A72417	prf. 1207243B	sp.RBSA_BACSU	pir.151116	sp.CYPA_MYCTU	sp:YQGP_BACSU
		ORF (bp)	993	180	528	1236	165	435	1413	438	1278	954	849	657	981	1023	759	816	561	687
45		Terminal (nt)	21065	21074	22124	23399	23615	24729	24885	26775	26822	28164	29117	30651	31677	32699	33457	33465	34899	35668
50		Initial (nt)	20073	21253	21597	22164	23779	24295	26297	26338	28099	29117	29965	29995	30697	31677	32699	1		34982
		SEO NO (a.a.)	3523	3524	3525	3526	3527	3528	3529	3530	3531	3532	3533	3534	3535	3536	3537	3538	3539	3540
<i>55</i>		SEQ NO.	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	, æ	39	40

EP 1 108 790 A2

																	<del></del>					
5		Function	ferric enterobactin transport system permease protein		-	vulnibactin utilization protein	hypothetical membrane protein	serine/threonine protein kinase	serine/threonine protein kinase	penicillin-binding protein	stage V sporulation protein E	phosphoprotein phosphatase	hypothetical protein	hypothetical protein					phenol 2-monooxygenase	succinate-semialdehyde dehydrogenase (NAD(P)+)	hypothetical protein	hypothetical membrane protein
			ferric er permea		ATPase	vulniba	hypothe	serine/t	serine/t	penicilli	stage V	hdsohd	hypothe	hypothe					phenol	succina dehydro	hypoth	hypothe
15		Matched length (a.a.)	332		253	260	95	648	486	492	375	469	155	526					117	490	242	262
20		Similarity (%)	70.5		81.8	52.7	72.6	68.7	59.1	66.7	65.6	70.8	66.5	38.8					63.3	78.2	9.73	64.1
		Identity (%)	40.4		51.8	26.2	40.0	40.6	31.7	33.5	31.2	44.1	38.7	23.6					29.9	46.7	27.3	29.0
25	(pən	au .	ဟ			viuB	osis	knB	pksC	Adc	VE	osis	osis	osis					ATCC	Q		ihi
30	Table 1 (continued)	Homologous gene	Escherichia coli K12 fepG		Vibrio cholerae viuC	Vibrio vulnificus MO6-24 viuB	Mycobacterium tuberculosis H37Rv Rv0011c	Mycobacterium leprae pknB	Streptomyces coelicolor pksC	Streptomyces griseus pbpA	Bacillus subtilis 168 spoVE	Mycobacterium tuberculosis H37Rv ppp	Mycobacterium tuberculosis H37Rv Rv0019c	Mycobacterium tuberculosis H37Rv Rv0020c					Trichosporon cutaneum ATCC 46490	Escherichia coli K12 gabD	Bacillus subtilis yrkH	Methanococcus jannaschii MJ0441
35 40	·	db Match	Sp:FEPG_ECOLI E		gp:VCU52150_9 \	5		SP. PKNB_MYCLE	gp:AF094711_1 S	gp:AF241575_1 S	)		pir.A70700	pir:B70700					sp:PH2M_TRICU	sp:GABD_ECOLI	sp:YRKH_BACSU [6	sp:Y441_METJA
		ORF (bp)	978	966	777	822	270	1938	1407	1422	1143	1353	462	864	147	720	219	471	954	1470	1467	789
45		Terminal (nt)	38198	36247	38978	39799	40189	40576	42513	43926	45347	46669	48024	48505	49455	49897	50754	99609	54008	51626	55546	55629
50		Initial (nt)	37221	37242	38202	38978	40458	42513	43919	45347	46489	48021	48485	49368	49601	50616	50972	51436	53055	53095	54080	56417
		SEQ NO.	3541	3542	3543	3544	3545	3546	3547	3548	3549	3550	3551	3552	3553	3554	3555	3556	3557	3558	3559	3560
55		NO SEQ	41	42	43	44	45	46	47	48	49	20	51	52	53	54	55	56	57	58	59	09

EP 1 108 790 A2

																	1			-	
	Function	hypothetical protein	hypothetical protein	hypothetical protein		hypothetical protein			magnesium and cobalt transport protein		chloride channel protein	required for NMN transport	phosphate starvation-induced protein-like protein				Mg(2+)/citrate complex secondary transporter	two-component system sensor histidine kinase		transcriptional regulator	D-isomer specific 2-hydroxyacid dehydrogenase
	Matched length (a.a.)	74	179	62		310			390		400	241	340				497	563		229	293
	Similarity (%)	74.3	70.4	83.9		50.7			59.5		64.8	53.1	0.09				68.8	9.09		63.3	73.7
	Identity (%)	40.5	36.3	53.2		26.8		_	29.5		30.0	24.1	29.1				42.3	27.2		33.2	43.3
lable 1 (confinued)	Homologous gene	Bacillus subtilis yrkF	Synechocystis sp. PCC6803 slr1261	Mycobacterium tuberculosis H37Rv Rv1766	-	Leishmania major L4768.11			Mycobacterium tuberculosis H37Rv Rv1239c corA		Zymomonas mobilis ZM4 clcb	Salmonella typhimurium pnuC	Mycobacterium tuberculosis H37Rv RV2368C				Bacillus subtilis citM	Escherichia coli K12 dpiB		Escherichia coli K12 criR	Corynebacterium glutamicum
	db Match	Sp.YRKF_BACSU	sp.YC61_SYNY3	pir.G70988		gp:LMFL4768_11			pir.F70952		gp:AF179611_12	SP:PNUC_SALTY	sp:PHOL_MYCTU				sp:CITM_BACSU	sp:DPIB_ECOLI		Sp. DPIA ECOLI	gp:AF134895_1
	ORF (bp)	291	591	174	855	840	711	1653	1119	447	1269	069	1122	132	384	765	1467	1653	570	654	912
	Terminal (nt)	56386	26680	57651	58941	59930	60662	62321	62390	63594	65458	65508	67972	68301	68251	69824	68720	72158	71474	72814	72817
	Initial (nt)	56676	57270	57478	58087	59091	59952	69909	63508	64040	64190	66197	66851	68170	68634	09069	70186	70506	72043	<del>ا</del> ــــــــــــــــــــــــــــــــــــ	<u> </u>
	SEQ NO.	3561	3562	3563	3564	3565	3566	3567	3568	3569	3570	3571	3572	3573	3574	3575	3576	3577	3578	3579	3580
	SEO NO.	) [6	62	63	64	65	99	29	68	69	70	17	72	73	74	75	76	17	78	62	8

5										lux protein				nt information	, ,		-	or	init or urease		
10		Function	hypothetical protein	biotin synthase	hypothetical protein	hypothetical protein		hypothetical protein	hypothetical protein	integral membrane efflux protein	creatinine deaminase			SIR2 gene family (silent information regulator)	triacylglycerol lipase	triacylglycerol lipase		transcriptional regulator	urease gammma subunit or urease structural protein	urease beta subunit	urease alpha subunit
15		Matched length (a.a.)	127	334	43	85		42	84	507	394			279	. 251	262		171	100	162	570
20		Similarity (%)	76.4	99.7	79.1	63.5		75.0	0.99	59.0	99.8			50.2	59.0	56.1		94.7	100.0	100.0	100.0
		Identity (%)	38.6	99.4	72.1	34.1		71.0	61.0	25.6	97.2			26.2	30.7	29.4		90.6	100.0	100.0	100.0
25	tinued)	ene	lor A3(2)	amicum	culosis	visiae		n Nigg	ae	e varS				visiae hst2	nes	nes		tamicum	tamicum	tamicum	tamicum
	Table 1 (continued)	Homologous gene	Streptomyces coelicolor A3(2) SCM2.03	Corynebacterium glutamicum bioB	Mycobacterium tuberculosis H37Rv Rv1590	Saccharomyces cerevisiae YKL084w		Chlamydia muridarum Nigg TC0129	Chlamydia pneumoniae	Streptomyces virginiae varS	Bacillus sp.			Saccharomyces cerevisiae hst2	Propionibacterium acnes	Propionibacterium acnes		Corynebacterium glutamicum ureR	Corynebacterium glutamicum ureA	Corynebacterium glutamicum ATCC 13032 ureB	Corynebacterium glutamicum ATCC 13032 ureC
35 40		db Match	gp:SCM2_3	sp:BIOB_CORGL	pir:H70542	sp:YKI4_YEAST	-	PIR:F81737	GSP:Y35814	prf:2512333A				sp:HST2_YEAST	prf.2316378A	prf.2316378A		gp:AB029154_1	gp:AB029154_2	gp:CGL251883_2	gp:CGL251883_3
		ORF (bp)	429	1002	237	339	117	141	273	1449	1245	306	615	924	972	900	888	513	300	486	1710
45		Terminal (nt)	74272	75491	75742	76035	76469	80613	81002	82120	83691	85098	85663	87241	87561	88545	90445	90461	91473	91988	93701
50		Initial (nt)	73844	74490	75506	75697	76353	80753	81274	83568	84935	85403	86277	86318	88532	89444	89558	90973	91174	91503	91992
		SEQ NO.	3581	3582	3583	3584	3585	3586	3587	3588	3589	3590	3591	3592	3593	3594	3595	3596	3597	3598	3599
55		SEQ NO.	81	82	83	84	85	8	87	88	68	8	91	92	93	94	95	96	97	86	66

EP 1 108 790 A2

																					$\overline{}$	_		
5		Function	urease accessory protein	urease accessory protein	urease accessory protein	urease accessory protein	epoxide hydrolase		valanimycin resistant protein			heat shock protein (hsp90-family)	AMP nucleosidase		acetolactate synthase large subunit		proline dehydrogenase/P5C dehydrogenase		aryl-alcohol dehydrogenase (NADP+)	pump protein (transport)	indole-3-acetyl-Asp hydrolase		hypothetical membrane protein	
15		Matched length (a.a.)	157	226	205	283	279		347			999	481		196	·	1297		338	513	352	•	106	
20		Similarity (%)	100.0	100.0	100.0	100.0	48.4		59.7			52.7	68.2		28.7		50.4		2.09	71.4	49.2		70.8	
		Identity (%)	100.0	100.0	100.0	100.0	21.2		26.5			23.8	41.0		29.6		25.8		30.2	36.5	23.0		35.9	
25	ntinued)	gene	itamicum	itamicum	itamicum	itamicum	bacter echA		iciens vimF			htpG	amn		1 APE2509		ium putA		sosporium	удаН	erans		yidH	
<i>30</i>	Table 1 (continued)	Homologous gene	Corynebacterium glutamicum ATCC 13032 ureE	Corynebacterium glutamicum ATCC 13032 ureF	Corynebacterium glutamicum ATCC 13032 ureG	Corynebacterium glutamicum ATCC 13032 ureD	Agrobacterium radiobacter echA		Streptomyces viridifaciens vlmF			Escherichia coli K12 htpG	Escherichia coli K12 amn		Aeropyrum pernix K1 APE2509		Salmonella typhimurium putA		Phanerochaete chrysosporium aad	Escherichia coli K12 ydaH	Enterobacter agglomerans		Escherichia coli K12 yidH	
40		db Match	gp:CGL251883_4	gp:CGL251883_5	gp:CGL251883_6	gp:CGL251883_7	prf:2318326B		gp:AF148322_1			sp:HTPG_ECOLI	sp. AMN_ECOLI		pir.E72483		sp:PUTA_SALTY		sp:AAD_PHACH	sp:YDAH_ECOLI	prf:2422424A		sp:YIDH_ECOLI	
		ORF (bp)	471	678	615	849	777	699	1152	675	2775	1824	1416	579	552	099	3456	114	945	1614	1332	669	366	315
45		Terminal (nt)	94199	94879	95513	96365	96368	98189	97319	100493	98808	101612	104909	105173	105841	106630	110890	111274	112318	114083	115478	114564	115943	116263
50		Initial (nt)	93729	94202	94899	95517	97144	97521	98470	99819	101582	103435	103494	105751	106392	107289	107435	111161	111374	112470	114147	115262	115578	115949
		SEO NO.	3600	3601	3602	3603	3604	3605	3606	3607	3608	3609	3610	3611	3612	3613	3614	3615	3616	3617	3618	3619	3620	3621
55		SEQ NO. (DNA)	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121

EP 1 108 790 A2

5		Function		transcriptional repressor	methylglyoxalase	hypothetical protein	mannitof dehydrogenase	D-arabinitol transporter		galactitol utilization operon repressor	xylulose kinase		pantoate-beta-alanine ligase	3-methyl-2-oxobutanoate hydroxymethyltransferase		DNA-3-methyladenine glycosylase		esterase		carbonate dehydratase	xylose operon repressor protein	macrolide efflux protein		
15		Matched length (a.a.)		258	126	162	497	435		260	451		279	271		188		270		201	357	418		
20		Similarity (%)		59.7	78.6	64.8	70.4	68.3	į	64.6	68.1		100.0	100.0		9.79		69.3		53.2	49.3	61.2		
		Identity (%)		29.5	57.9	37.0	43.5	30.3		27.3	45.0		100.0	100.0		42.0		39.3		30.9	24.1	21.1		
25	inued)	ene		ciens		ulosis	cens mtlD	dalT		atR	sus xylB		amicum	amicum		าลg		bacterium		ıophila	yIR	1214		
30	Table 1 (continued)	Homologous gene		Agrobacterium tumefaciens accR	Bacillus subtilis yurT	Mycobacterium tuberculosis H37Rv Rv1276c	Pseudomonas fluorescens mtlD	Klebsiella pneumoniae dalT		Escherichia coli K12 gatR	Streptomyces rubiginosus xylB		Corynebacterium glutamicum ATCC 13032 panC	Corynebacterium glutamicum ATCC 13032 panB		Arabidopsis thaliana mag		Petroleum-degrading bacterium HD-1 hde	,	Methanosarcina thermophila	Bacillus subtilis W23 xylR	Lactococcus lactis mef214		
35 40		db Match		sp:ACCR_AGRTU	pir.C70019	sp:YC76_MYCTU	prf.2309180A	prf.2321326A		sp:GATR_ECOLI	sp:XYLB_STRRU		gp:CGPAN_2	gp:CGPAN_1		Sp:3MG_ARATH		gp:AB029896_1		sp:CAH_METTE	sp:XYLR_BACSU	gp:LLLPK214_12		
		ORF (bp)	2052	780	390	510	1509	1335	189	837	1419	822	837	813	951	630	654	924	627	558	1143	1272	804	444
45		Terminal (nt)	116548	118810	120410	120413	120951	122507	124030	124966	126350	127992	126353	127192	128099	129489	130798	130815	132424	132981	132971	134207	135518	136122
50		Initial (nt)	118599	119589	120021	120922	122459	123841	123842	124130	124932	127171	127189	128004	129049	130118	130145	131738	131798	132424	134113	135478	136321	136565
		SEQ NO. (a.a.)	3622	3623	3624	3625	3626	3627	3628	3629	3630	3631	3632	3633	3634	3635	3636	3637	3638	3639	3640	3641	3642	3643
		Q O S	22	23	24	25	92	27	28	29	8	31	32	33	34	35	36	37	38	39	40	41	42	43

EP 1 108 790 A2

		_													$\overline{}$					-				· · · · ·			$\neg$
5								ne protein				itive protein	ne protein				ne protein			ase		e Je	secific for	e glycosylase	ي.		sis enzyme
10			Function				cellulose synthase	hypothetical membrane protein				chloramphenicol sensitive protein	hypothetical membrane protein			transport protein	hypothetical membrane protein			ATP-dependent helicase		nodulation protein	DNA repair system specific for alkylated DNA	DNA-3-methyladenine glycosylase	threonine efflux protein	hypothetical protein	doxorubicin biosynthesis enzyme
15			Matched length (a.a.)				420	593				303	198			361	248			829		188	219	166	217	55	284
20			Similarity (%)				51.2	51.8				60.7	59.1			62.3	70.2			64.3		0.99	60.7	65.1	61.3	72.7	52.1
			Identity (%)				24.3	25.1			-0-	34.7	30.3			32.4	34.7			33.8		40.4	34.7	39.8	34.1	50.9	31.0
25	:	ntinued)	gene				faciens celA	evisiae				ginosa rarD	yadS			abrB	2 yfcA			2 hrpB		sarum bv. IJI nodL	'3#1 alkB	2 tag	2 rhtC	A	etius dnrV
30		Table 1 (continued)	Homologous gene				Agrobacterium tumefaciens celA	Saccharomyces cerevisiae YDR420W hkr1				Pseudomonas aeruginosa rarD	Escherichia coli K12 yadS			Escherichia coli K12 abrB	Escherichia coli K12 yfcA			Escherichia coli K12 hrpB		Rhizobium leguminosarum bv. viciae plasmid pRL1JI nodL	Escherichia coli 0373#1 alkB	Escherichia coli K12 tag	Escherichia coli K12 rhtC	Bacillus subtilis yaaA	Streptomyces peucetius dnrV
<i>35</i> <i>40</i>			db Match				pir.139714	sp.HKR1_YEAST				sp:RARD_PSEAE	sp:YADS_ECOLI			sp. ABRB_ECOLI	sp:YFCA_ECOLI			sp:HRPB_ECOLI		sp:NODL_RHILV	SP.ALKB_ECOLI	sp:3MG1_ECOLI	sp:RHTC_ECOLI	sp:YAAA_BACSU	prf.2510326B
			ORF (bp)	1941	1539	636	1461	1731	621	1065	756	879	717	333	1659	1137	798	624	405	2388	315	675	069	525	8/9	291	852
45			Terminal (nt)	138744	140329	139226	141789	143526	143075	144639	145480	145518	147238	147570	149780	149794	152369	150966	152814	153226	156167	156147	157537	158138	158831	159159	160013
50			Initial (nt)	136804	138791	139861	140329	141796	142455	143575	144725	146396	146522	147238	148122	150930	151572	151589	152410	155613	155853	156821	156848	157614	158154		159162
			SEQ NO.	3644	3645	3646	3647	3648	3649	3650	3651	3652	3653	3654	3655	3656	3657	3658	3659	3660	3661	3662	3663	3664	3665	3666	3667
<i>55</i>			SEQ NO.	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167

EP 1 108 790 A2

		_												—,					ш,						$\overline{}$
5			on								lopeptidase 1		ator, GntR family ive regulator	ohydrate kinase	-	semialdehyde	ism	ism	n protein	drogenase	ism	ımp of sistance			
10			Function	methyltransferase				ribonuclease			neprilysin-like metallopeptidase		transcriptional regulator, GntR family or fatty acyl-responsive regulator	fructokinase or carbohydrate kinase	hypothetical protein	methylmalonic acid semialdehyde dehydrogenase	myo-inositol catabolism	myo-inositol catabolism	rhizopine catabolism protein	myo-inositol 2-dehydrogenase	myo-inositol catabolism	metabolite export pump of tetracenomycin C resistance		oxidoreductase	
15			Matched length (a.a.)	104	•			118			722		238	332	296	498	268	286	290	335	287	457		354	
20			Similarity (%)	56.7				76.3			57.2		65.6	63.0	80.7	86.1	58.2	8.69	51.0	72.2	72.1	61.5		65.5	
	•		Identity (%)	35.6				41.5			28.5		29.8	28.6	52.7	61.0	33.2	41.0	29.7	39.1	44.6	30.9		31.1	
25		tinued)	ene	s pombe	٠			s MC58					farR		ilor A3(2)	ilor msdA			၁၁၀	r iolG		scens tcmA			
30		Table 1 (continued)	Homologous gene	Schizosaccharomyces pombe SPAC1250.04c				Neisseria meningitidis MC58 NMB0662			Mus musculus nl1		Escherichia coli K12 farR	Beta vulgaris	Streptomyces coelicolor A3(2) SC8F11.03c	Streptomyces coelicolor msdA	Bacillus subtilis iolB	Bacillus subtilis iolD	Rhizobium meliloti mocC	Bacillus subtilis idh or iolG	Bacillus subtilis iolH	Streptomyces glaucescens tcmA		Bacillus subtilis yvaA	
35 40			db Match	gp:SPAC1250_3				gp:AE002420_13			gp:AF176569_1		sp.FARR_ECOLI	pir.T14544	gp:SC8F11_3	prf.2204281A	sp:IOLB_BACSU	sp:IOLD_BACSU	sp:MOCC_RHIME	sp:MI2D_BACSU	sp:IOLH_BACSU	sp:TCMA_STRGA		sp:YVAA_BACSU	
			ORF (bp)	342	930	657	933	405	639	741	2067	963	759	1017	921	1512	888	1728	954	101	870	1374	621	1023	456
45		٠	Terminal (nt)	160370	161360	162352	161363	162867	163603	166457	163689	167419	167837	169991	170916	172444	173355	175275	176272	177318	178203	179658	178461	180711	181297
50			Initial (nt)	160029	160431	161696	162295	162463	162965	165717	165755	166457	168595	168975	169996	170933	172468	173548	175319	176308	177334	178285	179081	179689	180842
			SEQ NO.	3668	3669	3670	3671	3672	3673	3674	3675	3676	3677	3678	3679	3680	3681	3682	3683	3684	3685	3686	3687	3688	3689
55			SEQ. NO. DNA)	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189

EP 1 108 790 A2

																							$\overline{}$	
. 5	-										hyltransferase		ıylase			orter			tidine kinase	ate e subunit	ate III subunit		-	
10		Function		regulatory protein	oxidoreductase	hypothetical protein		cold shock protein			caffeoyl-CoA 3-O-methyltransferase		glucose-resistance amylase regulator regulator			D-xylose proton symporter		transposase (ISCg2)	signal-transducing histidine kinase	glutamine 2-oxoglutarate aminotransferase large subunit	glutamine 2-oxoglutarate aminotransferase small subunit		hypothetical protein	
15		Matched length (a.a.)		331	442	303		64			134		338			458		401	145	1510	506		496	
20	ļ	Similarity (%)		61.9	52.5	64.7		92.2			58.2		62.1			70.5		100.0	60.7	100.0	99.8		72.8	
		identity (%)		32.0	24.4	33.7		70.3			30.6		28.7			36.0		100.0	27.6	99.9	99.4		44.6	
25	ontinued)	s gene		uli cebR	234 y4hM	_		icolor A3(2)					Ρφ			s xylT		glutamicum	fixL	glutamicum	glutamicum		berculosis	
30	Table 1, (continued)	Homologous gene		Streptomyces reticuli cebR	Rhizobium sp. NGR234 y4hM	Bacillus subtilis yfiH		Streptomyces coelicolor A3(2) csp			Stellaria longipes		Bacillus subtilis ccpA			Lactobacillus brevis xyIT		Corynebacterium glutamicum ATCC 13032 tnp	Rhizobium meliloti fixl	Corynebacterium glutamicum gltB	Corynebacterium glutamicum gltD		Mycobacterium tuberculosis H37Rv Rv3698	
35 40		db Match		gp:SRE9798_1	Sp.Y4HM_RHISN	sp:YFIH_BACSU		sp:CSP_ARTGO			prf:2113413A		sp.ccPA_BAcsU			sp:XYLT_LACBR		gp:AF189147_1	Sp:FIXL_RHIME	gp:AB024708_1	gp:AB024708_2		pir.C70793	
		ORF (bp)	384	993 94	1233 sp	1011 SF	429	201 sp	534	306	414 PI	426	ls 066	402	240	1473 s	300	1203 g	435 s	4530 g	1518 g	240	1485 p	369
45		Terminal (nt)	181647	181687	184051	1	185642	186708	187302	187607	188100	188300	188747	190321	190389	190703	192949	194464	194604	199769	201289	201341	201760	205956
50		Initial (nt)	181264	182679	182819	184077	185214	186508	186769	187302	187687	188725	189736	189920	190628	192175	193248	193262	195038		199772	201580	-1	205588
,		SEQ NO.	3690	3691	3692	3693	3694	3695	3696	3697	3698	3699	3700	3701	3702	3703	3704	3705	3706	3707	3708	3709	3710	3711
<i>55</i>		SEQ NO. (DNA)	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211

EP 1 108 790 A2

																					_
5		Function		arabinosyl transferase	hypothetical membrane protein	acetoacetyl CoA reductase	oxidoreductase				proteophosphoglycan	hypothetical protein		hypothetical protein	rhamnosyl transferase		hypothetical protein	O-antigen export system ATP- binding protein	O-antigen export system permease protein	hypothetical protein	NADPH quinone oxidoreductase
15		Matched length (a.a.)		1122	651	223	464				350	124		206	302		214	236	262	416	302
20		Similarity (%)		70.6	66.1	56.5	85.1				57.4	83.9		73.8	79.1		55.1	78.4	75.6	63.0	71.5
		Identity (%)		39.8	35.0	31.4	0.99				24.3	60.5		43.2	63.6		31.3	47.0	31.3	36.5	41.1
30	Table 1 (continued)	Homologous gene		Mycobacterium avium embB	Mycobacterium tuberculosis H37Rv Rv3792	Pseudomonas sp. phbB	Mycobacterium tuberculosis H37Rv Rv3790				Leishmania major ppg1	Mycobacterium tuberculosis H37Rv Rv3789		Mycobacterium tuberculosis H37Rv Rv1864c	Mycobacterium tuberculosis H37Rv Rv3782 rfbE		Agrobacterium tumefaciens plasmid pTi-SAKURA tiorf100	Yersinia enterocolitica rfbE	Yersinia enterocolitica rfbD	Mycobacterium tuberculosis H37Rv Rv3778c	ens pig3
	Tab	Horr		Mycobacter	Mycobacterium H37Rv Rv3792	Pseudomor	Mycobacter H37Rv Rv3				Leishmania	Mycobacter H37Rv Rv3		Mycobacterium to H37Rv Rv1864c	Mycobacterium tube H37Rv Rv3782 rfbE		Agrobacter plasmid pT	Yersinia en	Yersinia er	Mycobacte H37Rv Rv	Homo sapiens pig3
35 40		db Match		prf.2224383C	pir:D70697	prf:2504279B	pir.B70697				gp:LMA243459_1	sp:Y0GN_MYCTU		pir:H70666	pir.B70696		gp:AB016260_100	sp:RFBE_YEREN	Sp.RFBD_YEREN	pir.F70695	gp:AF010309_1
		ORF (bp)	318	3471 p	1983 р	759	1464 p	234	507	453	1002	396	402	633	939	342	297	789	804	1173	954
45		Terminal (nt)	206385	203541	207007	209210	209992	211535	212283	212735	213657	214107	214522	215159	215162	216605	216116	217141	217943	220151	220154
50		Initial (nt)	206068	207011	208989	209968	211455	211768	211777	212283	<del> </del>	213712	214121	214527	216100	216264	216712	217929	218746	218979	221107
	,	SEQ NO. (a.a.)	3712	3713	3714	3715	3716	3717	3718	3719	3720	3721	3722	3723	3724	3725	3726	3727	3728	3729	3730
55		SEQ NO. (DNA)	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

	Function		probable electron transfer protein	amino acid carrier protein		molybdopterin biosynthesis protein moeB (sulfurylase)	molybdopterin synthase, large subunit	molybdenum cofactor biosynthesis protein CB	co-factor synthesis protein	molybdopterin co-factor synthesis protein	hypothetical membrane protein	molybdate-binding periplasmic protein	molybdopterin converting factor subunit 1	maltose transport protein	hypothetical membrane protein	histidinol-phosphate aminotransferase			
	Matched length (a.a.)		78	475		368	150	158	154	377	227	256	96	365	121	330			
	Similarity (%)		51.0	75.8		70.1	75.3	63.3	84.4	58.6	70.5	68.0	70.8	8.09	76.9	65.8			
	Identity (%)		35.0	46.7		43.8	44.7	33.5	61.7	34.5	44.1	34.0	37.5	34.3	36.4	37.3			
Table 1 (continued)	Homologous gene		Mycobacterium tuberculosis H37Rv Rv3571	Bacilius subtilis alsT		Synechococcus sp. PCC 7942 moeB	Arthrobacter nicotinovorans moaE	Synechococcus sp. PCC 7942 moaCB	Arthrobacter nicotinovorans moaC	Arthrobacter nicotinovorans moeA	Arthrobacter nicotinovorans modB	Arthrobacter nicotinovorans modA	Mycobacterium tuberculosis H37Rv moaD2	Thermococcus litoralis malK	Streptomyces coelicolor A3(2) ORF3	Zymomonas mobilis hisC			
	db Match		PIR:A70606	sp.ALST_BACSU		gp:SYPCCMOEB_	prf.2403296D	sp:MOCB_SYNP7	prf.2403296C	gp:ANY10817_2	prf.2403296F	prf:2403296E	pir.D70816	prf.2518354A	sp:YPT3_STRCO	sp:HIS8_ZYMMO			
	ORF (bp)	582	. 297	1476	606	1083	456	471	468	1185	723	804	321	912	420	1023	906	294	120
	Terminal (nt)	221131	222207	222210	225244	225242	226312	226760	227218	227703	228891	229711	230928	230931	231848	232260	234818	234910	235409
	Initial (nt)	221712	221911	223685	224336	226324	226767	227230	227685	228887	229613	230514	230608	231842	232267	233282	233913	235203	235290
	SEQ NO. (a.a.)	3731	3732	3733	3734	3735	3736	3737	3738	3739	3740	3741	3742	3743	3744	3745	3746	3747	3748
	SEQ NO. (DNA)	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248

queuine tRNA-ribosyltransferase hypothetical membrane protein Na/dicarboxylate cotransporter membrane transport protein magnesium ion transporter glutamyl-tRNA synthetase nitrogen fixation protein alcohol dehydrogenase Function hypothetical protein putrescine oxidase transcription factor ABC transporter oxidoreductase transposase Matched length (a.a) Similarity 68.0 49.6 63.3 55.0 73.8 62.1 S 45.7 57.1 38.1 70.1 59. |dentity 41.3 24.3 34.8 33.2 34.2 29.4 34.0 21.5 30.9 48.8 20.7 45. 28. Streptomyces glaucescens strW Pseudomonas syringae tnpA Mycobacterium tuberculosis H37Rv Rv0507 mmpL2 Fable 1 (continued) Mycobacterium tuberculosis H37Rv tyrA Mycobacterium tuberculosis H37Rv Rv3753c Bacillus stearothermophilus DSM 2334 adh Bradyrhizobium japonicum Hamologous gene Borrelia burgdorferi mgtE Micrococcus rubens puo Brucella abortus oxyR Bacillus subtilis ypdP Zymomonas mobilis Bacillus subtilis gltX Xenopus faevis gp:PSESTBCBAD\_ sp:YPDP\_BACSU 2403 | sp:YV34\_MYCTU sp:ADH2\_BACST 1263 Sp.TGT\_ZYMMO gp:BAU81286\_1 sp:SYE\_BACSU sp:PUO\_MICRU gp:RHBNFXP db Match prf:2305239A prf.2320140A pir.S65588 pir.C70800 pir.B70800 ORF (bp) Terminal £ <u>E</u> NO. (a.a.) SEQ NO. 

` *25* 

EP 1 108 790 A2

5	Function	aspartate transaminase		DNA polymerase III holoenzyme tau subunit		hypothetical protein	recombination protein	cobyric acid synthase	UDP-N-acetylmuramyl tripeptide synthetase	DNA polymerase III epsilon chain	hypothetical membrane protein	aspartate kinase alpha chain			extracytoplasmic function alternative sigma factor	vegetative catalase			leucine-responsive regulatory protein	branched-chain amino acid transport
15	Matched length (a.a.)	432		642		101	214	248	444	346	270	421			189	492			143	203
20	Similarity (%)	100.0		53.1		74.3	72.4	61.7	9.09	55.2	100.0	8.66			63.5	76.4			72.0	68.0
	Identity (%)	98.6		31.6		41.6	42.5	38.3	31.3	25.7	100.0	99.5			31.2	52.9			37.1	30.5
25 Table 1 (continued)	ans gene	tofermentum		hilus dnaX		ıaK	cR	llis cobQ	ilis murC	berculosis	glutamicum avum) ATCC	glutamicum			negmatis sígE	atA			oniae Irp	A1 azlC
Table 1 (	Homologous gene	Brevibacterium lactofermentum aspC		Thermus thermophilus dnaX		Bacillus subtilis yaaK	Bacillus subtilis recR	Heliobacillus mobilis cobQ	Heliobacillus mobilis murC	Mycobacterium tuberculosis H37Rv dnaQ	Corynebacterium glutamicum (Brevibacterium flavum) ATCC 13032 orfX	Corynebacterium glutamicum lysC-alpha			Mycobacterium smegmatis sigE	Bacillus subtilis katA			Klebsiella pneumoniae Irp	Bacillus subtilis 1A1 azlC
35	db Match	gsp:W69554		gp.AF025391_1		Sp:YAAK_BACSU E	sp. RECR_BACSU E	prf:2503462B		pir.H70794	sp:YLEU_CORGL (	sp.AKAB_CORGL			prf.2312309A	sp.CATV_BACSU			sp:LRP_KLEPN	sp:AZLC_BACSU
40	ORF (bp)	1296 gsp:	630	2325 gp.A	717	309 sp:\	654 sp:F	750 prf.	1269 prf.2	1080 pir.t	867 sp:\	1263 sp:/	1053	1434	579 prf.:	1506 sp.(	342	291	462 sp:l	753 sp.,
45	Terminal Of (b)	257894 12	258529 6:	260875 23	258596 7	261295 3	262055 6	262546 7	<del>                                     </del>	264599 10	268258 8	270633 12	269524 10	273194 14	273542 5	275871 19	276232 3	275957 2	276302 4	277581 7
50	Initial (nt)	256599	257900	258551	259312	260987	261402	263295	264566	265678	269124	269371	270576	271761	274120	274366	275891	276247	276763	276829
	SEQ NO.	3771	3772	3773	3774	3775	3776	3777	3778	3779	3780	3781	3782	3783	3784	3785	3786	3787	3788	3789
55	SEQ NO. (DNA)	27.1	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289

EP 1 108 790 A2

5	Function			metalioregulatory protein	arsenic oxyanion-translocation pump membrane subunit	arsenate reductase				Na+/H+ antiporter or multiple resistance and pH regulation related protein D	Na+/H+ antiporter	Na+/H+ antiporter or multiple resistance and pH regulation related protein A				transcriptional activator	two-component system sensor histidine kinase	alkaline phosphatase		phosphoesterase	hypothetical protein
!	-			met	arse	arse	_	_		Na+ resi prot	Na+	Na+ resi prof				tran	two	alk		e e	hyp
15	Matched length (a.a.)			6	341	119				503	119	824				223	521	180		307	149
20	Similarity (%)			68.9	84.2	68.9				70.4	70.6	64.3				70.4	8.85	0.09		54.7	71.8
	Identity (%)			34.4	52.2	31.1				32.4	37.0	34.1				38.6	26.7	28.3		26.1	37.6
25 (penul)	ene			4 arsR	4 arsB	us arsC				лгрО	us mnhC	пгрА				s CH34	culosis	G1363 apl			
S S Table 1 (continued)	Homologous gene			Sinorhizobium sp. As4 arsR	Sinorhizobium sp. As4 arsB	Staphylococcus xylosus arsC				Bacillus firmus OF4 mrpD	Staphylococcus aureus mnhC	Bacillus firmus OF4 mrpA		,		Alcaligenes eutrophus CH34 czcR	Mycobacterium tuberculosis mtrB	Lactococcus lactis MG1363 apl		Baciltus subtilis ykuE	Bacillus subtilis yqeY
35 40	db Match			gp:AF178758_1	gp:AF178758_2	sp.ARSC_STAXY				gp:AF097740_4	prf:2504285D	-,		·		sp:CZCR_ALCEU	pri:2214304B	sp:APL_LACLA		pir.B69865	sp:YQEY_BACSU
	ORF (bp)	324	315	345 gg	1080 gj	387 sp	318	270	453	1530 91	381 pi	2886 9	1485	603	864	999	1467 pi	603 s	561	915 p	453 s
45	Terminal (nt)	277904	277987	278388	279893	280279	280349	280670	280949	281404	282937	283317	287857	287059	287966	289131	289777	292417	291273	292597	293991
50	Initial (nt)	277581	278301	278732	278814	279893	280666	280939	281401	282933	283317	286202	286373		288829		291243	291815	291833	293511	293539
	SEQ NO. (a.a.)	3790	3791	3792	3793	3794	3795	3796	3797	3798	3799	3800	3801	3802	3803	3804	3805	3806	3807	3808	3809
<i>55</i>	SEQ NO (DNA)	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309

Table 1 (continued)  Table 1 (continued)  Homologous gene (%) (%) (aa)  Mycobacterium leprae pon 1 48.3 77.1 782 (aa)  Mycobacterium tuberculosis 65.1 89.9 149 (ab)  Streptomyces coelicolor A3(2) 84.0 96.0 50 (aa)  Streptomyces coelicolor A3(2) 84.0 96.0 50 (ab)  Streptomyces coelicolor A3(2) 84.0 96.0 50 (ab)  Mycobacterium tuberculosis 65.1 89.9 149 (ab)  CSU Bacillus subtilis IcfA 37.3 68.9 440 (ab)  SCH17.10 Escherichia coli K12 shiA 37.3 68.9 440 (ab)  CSU Bacillus subtilis IcfA 31.1 59.9 534 (ab)  SCH2A.28 (ab)  SCH2A.2	ſ			Т		Т		T	T	$\Box$	$\neg$						T	T	$\Box$		$\neg \uparrow$	$\neg$	
Table 1 (continued)  Table 1 (continued)  Homologous gene (%) (%) (%) (aa) (aa)  Mycobacterium leprae pon 1 48.3 77.1 782 g  Streptomyces coelicolor A3(2) 40.9 63.4 71 r  Whigh Mycobacterium tuberculosis 65.1 89.9 149 r  Mycobacterium tuberculosis 65.1 89.9 149 r  HJTRV RV3678c  CSU Bacillus subtilis lcfA 31.1 59.9 65.4 127 f  SCH47.10c  CSU Bacillus subtilis lcfA 31.1 59.9 65.4 127 f  SCH47.10c  CSU Bacillus subtilis lcfA 31.1 59.9 65.4 127 f  SCH428c  CSU Bacillus subtilis lcfA 31.1 59.9 65.7 153 f  HILV Rhizobium leguminosarum nodN 45.8 72.6 153 f  Mycobacterium tuberculosis 30.9 65.7 207 r  Whycobacterium tuberculosis 24.5 77.1 240 Mycobacterium tuberculosis 24.5 77.1 240 241 241 241 241 241 241 241 241 241 241			Function	enicillin-binding BP1)	y protein		cal protein	tional regulator	s transport protein		in-fatty-acid-CoA ligase	tional regulator	n-(acyl-carrier-protein) e	e synthetase	ain acyl CoA oxidase	on protein	Ð			ceptor protein		t N-glycosylase/AP lyase	cytochrome c biogenesis protein
Table 1 (continued)  Table 1 (continued)  Homologous gene (%) (%) (%) (%)  Mycobacterium leprae pon1 48.3 77.1  Streptomyces coelicolor A3(2) 40.9 63.4  Whigh Mycobacterium tuberculosis 65.1 89.9  H37Rv Rv3678c  CSU Bacillus subtilis 16/A 31.1 59.9  Streptomyces coelicolor A3(2) 33.9 65.4  Streptomyces coelicolor A3(2) 33.9 65.4  Streptomyces coelicolor A3(2) 33.9 65.4  H37Rv Rv3677c  CSU Bacillus subtilis fabG 41.0 72.5  HILV Rhizobium leguminosarum nodN 45.8 72.6  Mycobacterium tuberculosis 41.2 72.4  H37Rv Rv3677c  CLU Micrococcus luteus pdg 57.5 77.1  Mycobacterium tuberculosis 24.5 65.7				class A p protein(P	regulator		hypotheti	transcript	shikimate		long-cha	transcrip(	3-oxoacy reductas	glutamin	short-cha	nodulatio	hydrolas			cAMP re		ultraviole	cytochro
Table 1 (continued)  Table 1 (continued)  Homologous gene (%)  Mycobacterium leprae pon1  Streptomyces coelicolor A3(2)  Aycobacterium tuberculosis  ESCH17.10c  Mycobacterium tuberculosis  Streptomyces coelicolor A3(2)  Aycobacterium tuberculosis  Arabidopsis thaliana atg6  Ara	5		Matched length (a.a.)	782	7.1		50	149	440		534	127	251	254	394	153	272			207		240	211
Table 1 (continued)  Homologous gene Mycobacterium leprae pon1 Streptomyces coelicolor A3(2) whiB Streptomyces coelicolor A3(2) whiB Streptomyces coelicolor A3(2) SCH17.10c Mycobacterium tuberculosis H37Rv Rv367Bc CSU Bacillus subtilis lcfA Streptomyces coelicolor A3(2) SCJ4.28c CSU Bacillus subtilis fabG IENI Emericella nidulans fluG Arabidopsis thaliana atg6 Arabidopsis thaliana atg6 Arabidopsis thaliana atg6 Arabidopsis thaliana atg6 Wycobacterium tuberculosis H37Rv Rv3677c Vibrio cholerae crp Wycobacterium tuberculosis	<b>)</b>		Similarity (%)	77.1	63.4		96.0	89.9	68.9		59.9	65.4	72.5	52.0	66.5	72.6	72.4			65.7		77.1	58.3
			Identity (%)	48.3	40.9		84.0	65.1	37.3		31.1	33.9	41.0	27.2	38.8	45.8	41.2			30.9		57.5	34.6
		linued)	ene	pon1	lor A3(2)		lor A3(2)	sisoins	shiA			lor A3(2)		nG	atg6	arum nodN	culosis					dg	culosis
	Table 1 (con	Table 1 (con	Homologous g	lycobacterium leprae	treptomyces coelico hiB		treptomyces coelico CH17,10c	lycobacterium tubero 137Rv Rv3678c	scherichia coli K12 s		acillus subtilis IcfA	treptomyces coelico	acillus subtilis fabG	mericella nidulans fl	rabidopsis thaliana	thizobium leguminos	Aycobacterium tuber 137Rv Rv3677c		•	fibrio cholerae crp		Aicrococcus luteus p	Mycobacterium tuber H37Rv Rv3673c
db Match  db Match  1.2209359A  1.2209359A  1.2209359A  1.2209359A  1.2209359A  1.2209359A  1.2209359A  1.2209359A  1.2209359A  1.270790  1.270790  1.270790  1.270790	<u>,</u>			Σ	ω ≥			ΣI				SS	<del> </del>		∢		≥ I			>			ZI
	,		db Match	prf:2209359A	pir:S20912		gp:SCH17_10	pir:G70790	sp:SHIA			gp:SCJ4_28	sp:FABG_BA(	sp:FLUG_EMI	prf.2512386A	sp:NODN_RH	pir.F70790			prf.2323349A		sp:UVEN_MIC	pir.B70790
ORF (bp) 2385 2385 339 192 153 609 609 1536 525 525 933 942 1173 705 681			ORF (bp)	2385	339	192	153	459	1353	609	1536	525	933	942	1194	471	843	1173	705	681	192	780	558
294004 294004 294004 297402 297622 298250 298250 298250 298250 298250 300695 300695 300695 300695 300695 300696 300700 305768 306700 306700 306700 306700	5		Terminal (nt)	294004	297402	297622	297783	298250	298332	300695	299726	301512	303099	304074	305263	305758	306700	305195	307504	306782	307727	308734	309302
	)		Initial (nt)	296388	297064	297431	297631	297792	299684	300087	301261	302036	302167	303133	304070	305288	305858	306367		307462	307918		308745
SEQ NO. (a.a.) 3810 3811 3812 3814 3815 3816 3817 3818 3822 3823 3823 3826 3827 3826 3827 3827 3827 3827 3827 3827 3827 3827			SEQ NO.	3810	3811	3812	3813	3814	3815	3816	3817	3818	3819	3820	3821	3822	3823	3824	3825	3826	3827	3828	3829
310 SEQ NO. (DNA) 310 311 311 311 311 311 311 311 311 311	<u>,</u>		SEQ NO. (DNA)	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329

-	.[				_		-			Ţ		-							-		$\neg$
5 10		Function	hypothetical protein	serine proteinase	epoxide hydrolase	hypothetical membrane protein	phosphoserine phosphatase	hypothetical protein	conjugal transfer region protein		hypothetical membrane protein	hypothetical protein	hypothetical protein				ATP-dependent RNA helicase	cold shock protein		DNA topoisomerase I	
15	i	Matched length (a.a.)	192	396	280	156	287	349	319		262	201	59				764	67		526	
20		Similarity (%)	56.3	71.0	52.1	77.6	65.5	60.2	66.5		63.7	64.2	84.8				66.1	88.1		81.6	
		Identity (%)	30.7	38.6	29.6	46.8	29.6	35.0	32.9		30.5	33.8	47.5				33.8	68.7		61.7	
25	Table 1 (continued)	ns gene	12 yeaB	berculosis	sp. C12 cEH	berculosis	prae serB	berculosis	89		iberculosis	ıberculosis	berculosis				prA	iformis S155		Iberculosis IopA	
30	Table 1 (	Homologous gene	Escherichia coli K12 yeaB	Mycobacterium tuberculosis H37Rv Rv3671c	Corynebacterium sp.	Mycobacterium tuberculosis H37Rv Rv3669	Mycobacterium leprae MTCY20G9.32C. serB	Mycobacterium tuberculosis H37Rv Rv3660c	Escherichia coli trbB		Mycobacterium tuberculosis H37Rv Rv3658c	Mycobacterium tuberculosis H37Rv Rv3657c	Mycobacterium tuberculosis H37Rv Rv3656c				Bacillus subtilis yprA	Arthrobacter globiformis SI55 csp		Mycobacterium tuberculosis H37Rv Rv3646c topA	
<i>35</i>			ш	ΣI	ပ	ΣI	ΣΣ	ΣI	ш		ΣI	ΣI	ΣI					۷۵		≥I	
40		db Match	sp.YEAB_ECOLI	pir:H70789	prf:2411250A	pir.F70789	pir:S72914	pir.E70788	pir.C44020		pir.C70788	pir.B70788	pir.A70788				sp:YPRA_BACSU	sp:CSP_ARTGO		pir:G70563	
		ORF (bp)	699	1191	993	549	996	1023	1023	615	816	546	198	318	414	345	2355	201	225	2988	711
<b>45</b>		Terminal (nt)	310038	311325	311899	312909	313625	316002	317132	316350	317893	318465	318689	319013	318545	319335	319336	322207	321992	325897	326614
50		Initial (nt)	309370	310135	312891	313457	314590	314980	316110	316964	317078	317920	318492	318696	318958	318991	321690	322007	322216	322910	325904
•		SEQ NO.	3830	3831	3832	3833	3834	3835	3836	3837	3838	3839	3840	3841	3842	3843	3844	3845	3846	3847	3848
55		SEQ NO.	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348

UDP-sugar hydrolase

586

54.4

26.1

Salmonella typhimurium ushA

2082 SP.USHA\_SALTY

345717 345814

367

162

				•									>	-	- ,	-					Г
5		Function	95	e III subunit		tein	tein	subunit ynthase C	e/xylosidase	<b>e</b>	dependent ehydrogenase		metallo-beta-lactamase superfamily	carrier-protein)	stant protein	,6-dehydratase	tein	ate mannose		rsynthetase	
10		Я	adenylate cyclase	DNA polymerase III subunit tau/gamma		hypothetical protein	hypothetical protein	ribosomal large subunit pseudouridine synthase C	beta-glucosidase/xylosidase	beta-glucosidase	NAD/mycothiol-dependent formaldehyde dehydrogenase		metallo-beta-lac	3-oxoacyl-(acyl-carrier-protein) reductase	valanimycin resistant protein	dTDP-glucose 4,6-dehydratase	hypothetical protein	dolichol phosphate mannose synthase		nucleotide sugar synthetase	
15		Matched length (a.a.)	263	423		144	172	314	558	101	362		160	251	415	320	108	230		260	
20		Similarity (%)	62.4	52.7		59.0	63.4	65.0	60.2	61.4	86.5		47.5	55.8	56.4	66.3	88.9	66.5		57.3	
		Identity (%)	32.7	25.3		32.6	39.0	43.6	34.8	38.6	9.99		32.5	25.9	26.3	33.8	59.3	33.9		25.8	
25	Table 1 (continued)	us gene	iaca B17R20	ıaX		ylicum uu033	odurans	12 rluC	emi D1 bgxA	nse salB	thanolica		hropolis orf5	12 fabG	difaciens vimF	срВ	berculosis	ınnaschii JAL-		12 yefJ	
30	Table 1 (	Homologous gene	Stigmatella aurantiaca B17R20 cyaB	Bacillus subtilis dnaX		Ureaplasma urealyticum uu033	Deinococcus radiodurans DR0202	Escherichia coli K12 rluC	Erwinia chrysanthemi D1 bgxA	Azospirillum irakense salB	Amycolatopsis methanolica		Rhodococcus erythropolis orf5	Escherichia coli K12 fabG	Streptomyces viridifaciens vlmF	Actinoplanes sp. acbB	Mycobacterium tuberculosis H37Rv Rv3632	Methanococcus jannaschii JAL- 1 MJ1222		Escherichia coli K12 yefJ	
40		db Match	CYAB_STIAU	sp:DP3X_BACSU E		o:AE002103_3	p:AE001882_8	sp:RLUC_ECOLI E	BGLX_ERWCH	5:AF090429_2	sp.FADH_AMYME A		sp:YTH5_RHOSN F	sp.FABG_ECOLI	3.AF148322_1	prf.2512357B A	r:A70562	sp:YC22_METJA		sp:YEFJ_ECOLI E	
		ORF (bp)	1041 sp	1257 s <sub>l</sub>	162	444 91	561 9	882   51	1644 sp	1989 gı	1104 s <sub>l</sub>	621	537 sl	ls 699	1230 gr	933 р	375 pi	759 sl	1029	1035 s	
45		Terminal (nt)	326695	329539	329909	330376	331533	332433	334562	334953	336112	335185	336748	337449	338768	339725	340195	340569	342375	343451	
50		Initial (nt)	327735	328283	329748	329933	330973	331552	332919	332965	335009	335805	336212	336781	337539	338793	340569	341327	341347	342417	
		SEQ NO.	3849	3850	3851	3852	3853	3854	3855	3856	3857	3858	3859	3860	3861	3862	3863	3864	3865	3866	
55		SEQ NO (DNA)	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	

EP 1 108 790 A2

	_																		<b>,</b>			
10		Function		NADP-dependent alcohol dehydrogenase	glucose-1-phosphate thymidylyltransferase	dTDP-4-keto-L-rhamnose reductase	dTDP-glucose 4,6-dehydratase	NADH dehydrogenase	Fe-regulated protein		hypothetical membrane protein	metallopeptidase	prolyl endopeptidase		hypothetical membrane protein	cell surface layer protein	autophosphorylating protein Tyr kinase	protein phosphatase		capsular polysaccharide biosynthesis	ORF 3	lipopolysaccharide biosynthesis / aminotransferase
15		Matched length (a.a.)		343	285	192	343	206	325		423	461	708		258	363	453	102		613	06	394
20		Similarity (%)		74.9	84.9	74.0	83.4	61.2	66.5		68.3	62.5	56.4		46.0	76.6	57.2	68.6		65.7	51.0	68.3
		Identity (%)		52.2	62.8	49.5	61.8	35.4	33.2		37.4	34.1	28.4		26.0	50.7	28.5	39.2		33.0	41.0	37.1
25	Table 1 (continued)	Homologous gene		luberculosis	um M32 rfbA	utans rmlC	utans XC rmlB	cus HB8 nox	aureus sirA		tuberculosis	oelicolor	capsulata		selicolor A3(2)	n ATCC 6872	hnsonii ptk	hnsonii ptp		Staphylococcus aureus M capD		ejuni wlaK
<i>30</i> <i>35</i>	Table 1	Homolog		Mycobacterium tuberculosis H37Rv adhC	Salmonella anatum M32 rfbA	Streptococcus mutans rmIC	Streptococcus mutans XC rmlB	Thermus aquaticus HB8 nox	Staphylococcus aureus sirA		Mycobacterium tuberculosis H37Rv Rv3630	Streptomyces coelicolor SC5F2A.19c	Sphingomonas capsulata		Streptomyces coelicolor A3(2)	Corynebacterium ammoniagenes ATCC 6872	Acinetobacter johnsonii ptk	Acinetobacter johnsonii ptp		Staphylococcus	Vibrio cholerae	Campylobacter jejuni wlaK
40		db Match		sp:ADH_MYCTU	sp:RFBA_SALAN	gp:D78182_5	SP:RMLB_STRMU	sp:NOX_THETH	prf:2510361A		sp:Y17M_MYCTU	gp:SC5F2A_19	prf.2502226A		gp:SCF43_2	gsp:W56155	prf.2404346B	prf.2404346A		sp:CAPD_STAAU	PRF:2109288X	prf.2423410L
		ORF (bp)	351	1059	855	1359	1131	579	945	639	1308	1380	2118	573	1092	1095	1434	603	984	1812	942	1155
45		Terminal (nt)	346110	346961	348098	348952	350313	351370	353637	353749	354599	355849	357237	359762	360814	362057	365257	365852	366838	368643	367701	369801
50		Initial (nt)	346460	348019	348952	350310	351443	351948	352693	354387	355906	357228	359354	360334	361905	363151	363824	365250	365855	366832	368642	368647
		SEQ NO. (a.a.)	3869	3870	3871	3872	3873	3874	3875	3876	3877	3878	3879	3880	3881	3882	3883	3884	3885	3886	3887	3888
55		SEQ NO. DNA)	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388

								_	1		<u> </u>		·	Ι	1	Τ			_		
5		c	tein	de	synthesis /	mine 1- se	cosamine				edneuce					rogenase					
10		Function	pilin glycosylation protein	capsular polysaccharide biosynthesis	lipopolysaccharide biosynthesis / export protein	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	UDP-N- acetylenolpyruvoylglucosamine reductase	sugar transferase	transposase		transposase (insertion sequence IS31831)	,	hypothetical protein	acetyltransferase	hypothetical protein B	UDP-glucose 6-dehydrogenase			glycosyl transferase	acetyltransferase	
15		Matched length (a.a.)	196	380	504	427	273	356	53		70		404	354	65	388			243	221	
20		Similarity (%)	75.0	69.2	69.8	64.6	68.5	57.3	79.3		94.3		57.4	60.2	53.0	89.7			65.0	62.0	
		Identity (%)	54.6	33.4	34.3	31.4	34.8	32.0	60.4		75.7		28.0	34.5	44.0	63.7			32.1	33.0	
<b>25</b>	ntinued)	gene	lis pglB	eus M capM	estris gumJ	e murA		.39x2	ntamicum		ıtamicum		rculosis	jinosa PAO1	ıtamicum				A	7 wbhH	
30	Table 1 (continued)	Homologous gene	Neisseria meningitidis pglB	Staphylococcus aureus M capM	Xanthomonas campestris gumJ	Enterobacter cloacae murA	Bacillus subtilis murB	Vibrio cholerae ORF39x2	Corynebacterium glutamicum		Corynebacterium glutamicum ATCC 31831		Mycobacterium tuberculosis H37Rv Rv1565c	Pseudomonas aeruginosa PAO1 psbC	Corynebacterium glutamicum	Escherichia coli ugd			Escherichia coli wbnA	Escherichia coli 0157 wbhH	
35		db Match	gp:AF014804_1	sp:CAPM_STAAU	pir:S67859	sp:MURA_ENTCL	sp:MURB_BACSU	gp:VCLPSS_9	prf.2211295A		pir:S43613		pir.G70539	gsp:W37352	PIR:S60890	sp:UDG8_ECOLI	2		gp:AF172324_3	gp:AB008676_13	
		ORF (bp)	612	1161	1491	1314	1005	1035	150	135	327	276	1170	993	231	1161	273	1209	822	645	195
45		Terminal (nt)	370405	371773	373419	374813	375837	376876	377832	378227	378511	378287	378668	379850	381495	383108	383496	383982	385374	387200	387463
50		Initial (nt)	369794	370613	371929	373500	374833	375842	377683	378093	378185	378562	379837	380842	381265	381948	383768	385190	386195	386556	387657
	_	SEQ NO. (a.a.)	3889	3890	3891	3892	3893	3894	3895	968£	3897	3898	3899	3900	3901	3902	3903	3904	3905	3906	3907
55		SEQ NO. (DNA)	389	390	391	392	393	394	395	396	397	398	399	400	401	405	403	404	405	406	407

	٠ (																				
	-		se						unit B	) = ·		- 11				-					
5		Function	dihydrolipoamide dehydrogenase	UTP-glucose-1-phosphate uridylyltransferase	regulatory protein	transcriptional regulator	cytochrome b subunit	succinate dehydrogenase flavoprotein	succinate dehydrogenase subunit						hypothetical protein	hypothetical protein			tetracenomycin C transcription repressor		transporter
15		Matched length (a.a.)	469	295	153	477	230	809	258						259	431			197		499
20		Similarity (%)	100.0	68.1	71.9	81.3	67.4	61.2	56.2						49.8	64.3			53.8		74.6
		Identity (%)	9.66	41.7	43.8	67.0	34.8	32.4	27.5						26.3	32.7			26.4		36.1
25	ontinued)	s gene	glutamicum	pestris	uginosa PAO1	erculosis	icolor A3(2)	Ψ	erans sdhB						icolor	2 yjiN			sueoseo		ae T#2717
30	Table 1 (continued)	Homologous gene	Corynebacterium g ATCC 13032 lpd	Xanthomonas campestris	Pseudomonas aeruginosa PAO1 orfX	Mycobacterium tuberculosis H37Rv Rv0465c	Streptomyces coelicalor A3(2) SCM10.12c	Bacillus subtilis sdhA	Paenibacillus macerans sdhB		٠				Streptomyces coelicolor SCC78.05	Escherichia coli K12 yjiN			Streptomyces glaucescens GLA.0 tcmR		Streptomyces fradiae T#2717 urdJ
35			0 ∢	×		ΣI	S S	<u> </u>							တ်တ						
40		db Match	gp:CGLPD_1	pir.JC4985	gp:PAU49666_2	pir.E70828	gp:SCM10_12	pir.A27763	gp.BMSDHCAB_4						gp:SCC78_5	sp:YJIN_ECOLI			sp:TCMR_STRGA		gp:AF164961_8
	,	ORF (bp)	1407	921	498	1422	771	1875	837	336	261	630	96	339	975	1251	420	303	829	204	1647
45		Terminal (nt)	389098	390168	390730	390787	393475	395513	396262	396650	396932	396411	397825	398222	397232	399579	400017	400341	401150	401253	402796
50		Initial (nt)	387692	389248	390233	392208	392705	393639	395426	396315	396672	397040	397730	397884	398206	398329	399598	400039	400473	401050	401150
		SEQ NO. (a.a.)	3908	3909	3910	3911	3912	3913	3914	3915	3916	3917	3918	3919	3920	3921	3922	3923	3924	3925	3926
55		SEQ NO. (DNA)	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426

EP 1 108 790 A2

5	Function	transporter	formyltetrahydrofolate deformylase	deoxyribose-phosphate aldolase			hypothetical protein	hypothetical protein		cation-transporting P-type ATPase B		glucan 1,4-alpha-glucosidase	hemin-binding periplasmic protein	ABC transporter	ABC transporter ATP-binding protein	hypothetical protein	hypothetical protein			
15	Matched length	1	286 fe	208 d			280 h	92 h		748 c		626 g	348 h	330 /	254 A	266 h	258 h			
20	Similarity (%)	74.6	72.7	74.0			53.6	85.9		75.3		56.1	83.6	90.3	85.0	56.4	61.6			
	Identity (%)	39.6	40.9	38.5			26.8	58.7		45.7		27.3	57.2	65.2	63.8	28.6	32.6	_		
25 (panuliuned) (continued)	us gene	Jiae T#2717	sp. P-1 purU	soC			vium GIR10	iberculosis		prae ctpB		erevisiae sta1	diphtheriae	diphtheriae	diphtheriae	elicolor C75A	elicolor C75A			
Table 1	Homologous gene	Streptomyces fradiae T#2717 urdJ	Corynebacterium sp.	Bacillus subtilis deoC			Mycobacterium avium GIR10 mav346	Mycobacterium tuberculosis H37Rv Rv0190		Mycobacterium leprae ctpB		Saccharomyces cerevisiae S288C YIR019C sta1	Corynebacterium diphtheriae hmuT	Corynebacterium diphtheriae hmuU	Corynebacterium diphtheriae hmuV	Streptomyces coelicolor C75A SCC75A.17c	Streptomyces coelicolor C75A SCC75A.17c			
<b>35</b>			<del>                                     </del>				ΣE	∑ï												
40	db Match	gp. AF164961_8	sp:PURU_CORSP	sp.DEOC_BACSU			prf:2413441K	pir.A70907		SP:CTPB_MYCLE		sp:AMYH_YEAST	gp:AF109162_1	gp:AF109162_2	gp:AF109162_3	gp:SCC75A_17	gp:SCC75A_17			
	ORF (bp)	1632	912	999	150	897	867	300	900	2265	450	1863	1077	1068	813	957	837	810	813	501
45	Terminal (nt)	404430	404508	406145	406161	405521	407416	407409	409145	407711	410027	412545	413633	414710	415526	416599	417439	417545	418441	419257
50	Initial (nt)	402799	405419	405480	406310	406417	406550	407708	408546	409975	410476	410683	412557	413643	414714	415643	416603	418354	419253	419757
	SEQ.	3927	3928	3929	3930	3931	3932	3933	3934	3935	3936	3937	3938	3939	3940	3941	3942	3943	3944	3945
55	SEQ	(DINA)	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445

EP 1 108 790 A2

																				,	
5			Function	UDP-N-acetylpyruvoylglucosamine reductase				long-chain-fatty-acidCoA ligase	transferase	phosphoglycerate mutase	two-component system sensor histidine kinase	two-component response regulator		ABC transporter ATP-binding protein	cytochrome P450	exopolyphosphatase	hypothetical membrane protein	pyrroline-5-carboxylate reductase	membrane glycoprotein	hypothetical protein	
15			Matched length (a.a.)	356				558	416	246	417	231		921	269	306	302	269	394	55	
20			Similarity (%)	58.4				68.1	58.7	84.2	74.8	90.9		60.7	6.99	8.73	57.3	100.0	52.0	94.6	
			Identity (%)	30.1				35.5	33.9	7.07	49.2	75.8		31.3	45.0	28.8	28.8	100.0	25.4	76.4	
25	7	iinea)		12 murB						or A3(2)	enX3	3CG		or A3(2)	losis	osa ppx	losis	micum	JRF71		
30	+ 	lable i (confinued)	Homologous gene	Escherichia coli RDD012 murB				Bacillus subtilis IcfA	Streptomyces coelicolor SC2G5.06	Streptomyces coelicolor A3(2) gpm	Mycobacterium bovis senX3	Mycobacterium bovis BCG regX3		Streptomyces coelicolor A3(2) SCE25.30	Mycobacterium tuberculosis H37Rv RV3121	Pseudomonas aeruginosa ppx	Mycobacterium tuberculosis H37Rv Rv0497	Corynebacterium glutamicum ATCC 17965 proC	Equine herpesvirus 1 ORF71	Mycobacterium leprae B2168_C1_172	
<i>35</i> <i>40</i>			db Match	gp:ECOMURBA_1				sp:LCFA_BACSU	gp:SC2G5_6	sp.PMGY_STRCO	prf:2404434A	prf:2404434B		gp:SCE25_30	sp:YV21_MYCTU	prf.2512277A	sp:YV23_MYCTU	sp.PROC_CORGL	gp:D88733_1	pir.S72921	
			ORF (bp)	1101	651	735	174	1704	1254 g	744 s	1239	969	879	2586 g	s 606	927	813 8	810	1122 g	198	219
45		,	Terminal (nt)	420885	421516	420309	422031	422090	425131	425920	427172	427867	429439	429438	432126	433988	434822	435695	433865	436137	436103
50			Initial (nt)	419785	420866	421043	421858	423793	423878	425177	425934	427172	428561	432023	433028	433062	434010	434886	434986	435940	436321
			SEQ NO. (a.a.)	3946	3947	3948	3949	3950	3951	3952	3953	3954	3955	3956	3957	3958	3959	3960	3961	3962	3963
			go≨	49	47	8	6	8	51	22	23	22	22	92	57	28	59	8	51	62	63

	_											
5		Function	hypothetical protein			phosphoserine phosphatase	hypothetical protein		glutamyl-tRNA reductase	hydroxymethylbilane synthase		
15	٠.	Matched length (a.a.)	29			296	74		455 (	308		
20		identity Similarity (%)	100.0			77.4	66.2		74.3	75.3		3
		identity (%)	89.7			51.0	40.5		44.4	50.7		7.20
25	intinued)	gene	olor			ae erB	erculosis		ae hemA	ae hem3b		aceticus
30	Table 1 (continued)	Homologous gene	Streptomyces coelicolor SCE68.25c			Mycobacterium leprae MTCY20G9.32C. serB	Mycobacterium tuberculosis H37Rv Rv0508		Mycobacterium leprae hemA	Mycobacterium leprae hem3b		Acinetobacter calcoaceticus
<i>35</i>		db Match	gp:SCE68_25			pir.S72914	sp:YV35_MYCTU		Sp. HEM1_MYCLE	pir.S72887		,
		ORF (bp)	66	192	618	1065	246	258	1389	906	372	0
45		Terminal (nt)	436561	436764	437850	436980	438424	438037	439904	440814	441591	, , , ,
50		Initial (nt)	436463	436573	437233	438044	438179	438294	438516	439909	441220	
		SEQ NO. (a.a.)	3964	3965	3966	3967	3968	3969	3970	3971	3972	
		0 8		1.0	100	_	m			l _	١.,	Ι.

Function	hypothetical protein			phosphoserine phosphatase	hypothetical protein		glutamyl-tRNA reductase	hydroxymethylbilane synthase		cat operon transcriptional regulator	shikimate transport protein	3-dehydroshikimate dehydratase	shikimate dehydrogenase		putrescine transport protein		iron(III)-transport system permease protein		periplasmic-iron-binding protein	uroporphyrin-III C-methyltransferase	
Matched length (a.a.)	29			296	74		455	808		321	417	608	282		363		829		347	486	
Similarity (%)	100.0			77.4	66.2		74.3	75.3		9'29	72.2	67.9	98.6		9.89		55.2		59.9	71.6	
Identity (%)	89.7			51.0	40.5		44.4	50.7		27.1	35.5	28.2	98.2		34.7		25.1		25.1	46.5	
Homologous gene	Streptomyces coelicolor SCE68.25c			Mycobacterium leprae MTCY20G9.32C. serB	Mycobacterium tuberculosis H37Rv Rv0508		Mycobacterium leprae hemA	Mycobacterium leprae hem3b		Acinetobacter calcoaceticus catM	Escherichia coli K12 shiA	Neurospora crassa qa4	Corynebacterium glutamicum ASO19 aroE		Escherichia coli K12 potG		Serratia marcescens sfuB		Brachyspira hyodysenteriae bitA	Mycobacterium leprae cysG	
db Match	gp:SCE68_25			pir.S72914	sp:YV35_MYCTU		sp.HEM1_MYCLE	pir.S72887		sp.CATM_ACICA	sp:SHIA_ECOLI	sp:3SHD_NEUCR	gp:AF124518_2		sp:POTG_ECOLI		sp:SFUB_SERMA		gp:SHU75349_1	pir:S72909	
ORF (bp)	66	192	618	1065	246	258	1389	906	372	882	1401	1854	849	273	1050	615	1644	1113	1059	1770	426
Terminal (nt)	436561	436764	437850	436980	438424	438037	439904	440814	441591	441601	444158	446038	447386	447398	448130	449100	449183	451961	450837	454430	454875
Initial (nt)	436463	436573	437233	438044	438179	438294	438516	439909	441220	442482	442758	444185	446538	447670	449179	449714	450826	450849	451895	452661	454450
SEQ NO. (a.a.)	3964	3965	3966	3967	3968	3969	3970	3971	3972	3973	3974	3975	3976	3977	3978	3979	3980	3981	3982	3983	3984
SEQ NO.	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484

					,	,					,			·							
5		Function	delta-aminolevulinic acid dehydratase			cation-transporting P-type ATPase B		uroporphyrinogen decarboxylase	protoporphyrinogen IX oxidase	glutamate-1-semialdehyde 2,1- aminomutase	phosphoglycerate mutase	hypothetical protein	cytochrome c-type biogenesis protein	hypothetical membrane protein	cytochrome c biogenesis protein		transcriptional regulator	Zn/Co transport repressor		hypothetical membrane protein	1,4-dihydroxy-2-naphthoate octaprenyltransferase
15		Matched length (a.a.)	337			858		364	464	425	161	208	245	533	338		144	06		82	301
20		Similarity (%)	83.1			56.5		7.97	59.9	83.5	62.7	71.2	85.3	76.0	77.8		69.4	72.2		78.1	61.5
		Identity (%)	8.09			27.4		55.0	28.0	61.7	28.0	44.7	53.5	50.7	44.1		38.9	31.1		39.0	33.6
25 30	Table 1 (continued)	Homologous gene	Streptomyces coelicolor A3(2) hemB			Mycobacterium leprae ctpB		Streptomyces coelicolor A3(2) hemE	Bacillus subtilis hemY	Mycobacterium leprae hemL	Escherichia coli K12 gpmB	Mycobacterium tuberculosis H37Rv Rv0526	Mycobacterium tuberculosis H37Rv ccsA	Mycobacterium tuberculosis H37Rv Rv0528	Mycobacterium tuberculosis H37Rv ccsB		Mycobacterium tuberculosis H37Rv Rv3678c pb5	Staphylococcus aureus zntR		Mycobacterium tuberculosis H37Rv Rv0531	Escherichia coli K12 menA
<i>35</i>		. ·	Streptom hemB			Mycobac		Streptom hemE	Bacillus s	Mycobac	Escheric	Mycobacterium H37Rv Rv0526	Mycobacterii H37Rv ccsA	Mycobacterium H37Rv Rv0528	Mycobacteriu H37Rv ccsB		Mycobaci H37Rv R	Staphyloc		Mycobacterium H37Rv Rv0531	Escherich
40		db Match	sp:HEM2_STRCO			SP:CTPB_MYCLE		sp.DCUP_STRCO	sp.PPOX_BACSU	sp:GSA_MYCLE	sp:PMG2_ECOL1	pir.A70545	pir:B70545	pir.C70545	pir:D70545	-	pir:G70790	prf:2420312A		pir.F70545	sp:MENA_ECOLI
		ORF (bp)	1017	582	510	2544	843	1074	1344	1311	909	621	792	1623	1011	801	471	357	300	333	894
45		Terminal (nt)	455983	456597	457150	459900	458583	461093	462455	463867	464472	465102	465909	467571	468658	470170	470654	470657	471121	471847	471915
50		Initial (nt)	454967	456016	456641	457357	459425	460020	461112	462557	463867	464482	465118	465949	467648	469370	470184	471013	471420	471515	472808
		SEQ NO. (a.a.)	3985	3986	3987	3988	3989	3990	3991	3992	3993	3994	3995	3996	3997	3998	3999	4000	4001	4002	4003
55		SEQ NO. (DNA)	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503

EP 1 108 790 A2

5	Function	glycosyl transferase	malonyl-CoA-decarboxylase	hypothetical membrane protein	ketoglutarate semialdehyde dehydrogenase	5-dehydro-4-deoxyglucarate dehydratase	als operon regulatory protein	hypothetical protein		2-pyrone-4,6-dicarboxylic acid				low-affinity inorganic phosphate transporter			naphthoate synthase	peptidase E	pterin-4a-carbinolamine dehydratase	muconate cycloisomerase
15	Matched length (a.a.)	238	421	139	929	303	293	94		267				410			293	202	77	335
20	Similarity (%)	62.6	51.5	65.5	76.0	9'5'	66.2	64.9		54.7				83.2			70.3	82.7	68.8	76.7
	Identity (%)	32.4	25.4	35.3	50.4	48.5	36.9	33.0		28.1				0.09			48.5	67.9	37.7	54.0
25 Table 1 (continued)	us gene	is wcgB	matB	(12 yajF	ıtida	tida KDGDH	68 alsR	berculosis		). LB126 fldB				berculosis			enB	odurans	/F5 phhB	berculosis enC
Table 1	Homologous gene	Bacteroides fragilis wcgB	Rhizobium trifolii matB	Escherichia coli K12 yajF	Pseudomonas putida	Pseudomonas putida KDGDH	Bacillus subtilis 168 alsR	Mycobacterium tuberculosis H37Rv Rv0543c		Sphingomonas sp.				Mycobacterium tuberculosis H37Rv pitA			Bacillus subtilis menB	Deinococcus radiodurans DR1070	Aquifex aeolicus VF5 phhB	Mycobacterium tuberculosis H37Rv Rv0553 menC
35					Ъ			ΣI		6.				ΣÏ					Ä	ΣÏ
40	db Match	gp:AF125164_6	prf:2423270B	sp:YQJF_ECOLI	pir.S27612	sp:KDGD_PSEPU	sp.ALSR_BACSU	pir.B70547		gp:SSP277295				pir.D70547			sp:MENB_BACSU	gp:AE001957_12	pir:C70304	pir.D70548
	ORF (bp)	864	1323	411	1560	948	879	315	444	750	417	378	261	1275	222	306	957	603	309	1014
45	Terminal (nt)	473811	473814	474997	475489	477048	478092	478989	480597	479452	480208	480624	481131	481394	483366	483637	484106	485986	485077	487014
50	Initial (nt)	472948	475136	475407	477048	477995	478970	479303	480154	480201	480624	481001	481391	482668	483587	483942	485062	485384	485385	486001
	SEQ NO. (a.a.)	4004	4005	4006	4007	4008	4009	4010	4011	4012	4013	4014	4015	4016	4017	4018	4019	4020	4021	4022
55	SEQ NO.	504	505	206	202	208	509	510	511	512	513	514	515	516	517	518	519	520	521	522

.

		T .	T		T		$\overline{}$		T						
	Function	2-oxoglutarate decarboxylase and 2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase	hypothetical membrane protein	alpha-D-mannose-alpha(1- 6)phosphatidyl myo-inositol monomannoside transferase	D-serine/D-alanine/glycine transporter	ubiquinone/menaquinone biosynthesis methyltransferase		oxidoreductase	heptaprenyl diphosphate synthase component II	preprotein translocase SecE subunit	transcriptional antiterminator protein	50S ribosomal protein L11	50S ribosomal protein L1	regulatory protein	4-aminobutyrate aminotransferase
	Matched length (a.a.)	909	148	408	447	237		412	316	111	318	145	236	564	443
	Similarity (%)	54.0	64.9	54.2	89.9	66.7		76.7	67.1	100.0	100.0	100.0	100.0	50.2	82.4
	Identity (%)	29.4	37.2	22.8	66.2	37.1		49.0	39.2	100.0	100.0	100.0	100.0	23.1	60.5
Table 1 (continued)	Homologous gene	Bacillus subtilis menD	Mycobacterium tuberculosis H37Rv Rv0556	Mycobacterium tuberculosis H37Rv pimB	Escherichia coli K12 cycA	Escherichia coli K12 ubiE		Mycobacterium tuberculosis H37Rv Rv0561c	Bacillus stearothermophilus ATCC 10149 hepT	Corynebacterium glutamicum ATCC 13032 secE	Corynebacterium glutamicum ATCC 13032 nusG	Corynebacterium glutamicum ATCC 13032 rplK	Corynebacterium glutamicum ATCC 13032 rpIA	Streptomyces coelicolor SC5H4.02	Mycobacterium tuberculosis H37Rv RV2589 gabT
	db Match	sp:MEND_BACSU	pir:G70548	pir.H70548	sp:CYCA_ECOLI	sp:UBIE_ECOLI		pir.D70549	sp:HEP2_BACST	gp:AF130462_2	gp:AF130462_3	gp:AF130462_4	gp:AF130462_5	gp:SC5H4_2	sp:GABT_MYCTU
	ORF (bp)	1629	441	1239	1359	069	699	1272	1050	333	954	435	708	1512	1344
	Terminal (nt)	488656	489100	490447	491938	492655	493583	492645	495110	497142	498327	499032	499869	499925	502920
	Initial (nt)	487028	488660	489209	490580	491966	492915	493916	494061	496810	497374	498598	499162	501436	501577
	SEQ NO. (a.a.)	4023	4024	4025	4026	4027	4028	4029	4030	4031	4032	4033	4034	4035	4036
	SEQ NO. (DNA)	523	524	525	526	527	528	529	530	531	532	533	534	535	929

EP 1 108 790 A2

5	Function	succinate-semialdehyde dehydrogenase (NAD(P)+)	novel two-component regulatory system	tyrosine-specific transport protein	cation-transporting ATPase G	hypothetical protein or dehydrogenase		50S ribosomal protein L10	50S ribosomal protein L7/L12		hypothelical membrane protein	DNA-directed RNA polymerase beta chain	DNA-directed RNA polymerase beta chain	hypothetical protein		DNA-binding protein	hypothetical protein
15	Matched length (a.a.)	461	150	447	615	468		170	130		283	1180	1332	169		232	215
20	Similarity (%)	71.8	38.0	49.9	64.4	66.2		84.7	89.2		55.5	90.4	88.7	52.0		63.8	57.7
	Identity (%)	40.8	32.0	25.5	33.2	40.2		52.9	72.3		25.8	75.4	72.9	39.0		39.2	29.3
os Table 1 (continued)	us gene	12 gabD	ense carR	12 0341#7	berculosis ctpG	ans P49		eus N2-3-11	perculosis IL		oerculosis	oerculosis oB	perculosis oC	erculosis		icolor A3(2)	erculosis
30 Table 1 (	Homologous gene	Escherichia coli K12 gabD	Azospirillum brasilense carR	Escherichia coli K12 o341#7 tyrP	Mycobacterium tuberculosis H37Rv RV1992C ctpG	Streptomyces lividans P49		Streptomyces griseus N2-3-11 rplJ	Mycobacterium tuberculosis H37Rv RV0652 rplL		Mycobacterium tuberculosis H37Rv Rv0227c	Mycobacterium tuberculosis H37Rv RV0667 rpoB	Mycobacterium tuberculosis H37Rv RV0668 rpoC	Mycobacterium tuberculosis H37Rv Jv0166c		Streptomyces coelicolor A3(2) SCJ9A, 15c	Mycobacterium tuberculosis H37Rv RV2908C
35	٠										ΣI						
40	db Match	sp:GABD_ECOLI	GP. ABCARRA_2	sp:TYRP_ECOLI	sp:CTPG_MYCTU	sp:P49_STRLI		sp:RL10_STRGR	sp:RL7_MYCTU		pir.A70962	sp:RPOB_MYCTU	sp:RPOC_MYCTU	GP:AF121004_1		gp:SCJ9A_15	sp:YT08_MYCTU
	ORF (bp)	1359	468	1191	1950	1413	603	513	384	138	972	3495	3999	582	180	780	798
45	Terminal (nt)	504283	503272	505569	507647	509081	509696	510510	510974	510989	512507	516407	520492	518696	520850	521644	521679
50	Initial (nt)	502925	503739	504379	505698	507669	509094	509998	510591	511126	511536	512913	516494	519277	520671	520865	522476
	SEQ NO.	4037	4038	4039	4040	4041	4042	4043	4044	4045	4046	4047	4048	4049	4050	4051	4052
55	SEQ NO. (DNA)	537	538	539	540	541	542	543	544	545	546	547	548	549	920	551	552

	-						$\top$			o'	tein	tein	<b>4</b>									
	Function	30S ribosomal protein S12	30S ribosomal protein S7	elongation factor G			lipoprotein			ferric enterobactin transport ATP- binding protein	ferric enterobactin transport protein	ferric enterobactin transport protein	butyryl-CoA:acetate coenzyme A transferase	30S ribosomal protein S10	50S ribosomal protein L3		50S ribosomal protein L4	50S ribosomal protein L23		50S ribosomal protein L2	30S ribosomal protein S19	
	_	308	30S	elon	_	_	od	_	-	ferri	ferri	ferri	buty tran	308	508	-	208	508	-	20S	308	_
	Matched length (a.a.)	121	154	709			44			258	329	335	145	101	212		212	98		280	92	
20	Similarity (%)	97.5	94.8	88.9			78.0			83.7	77.8	90.0	79.3	0.66	9.68		90.1	90.6		92.9	98.9	
	Identity (%)	6.06	81.8	71.7			56.0			56.2	45.6	48.1	56.6	84.2	66.5		71.2	74.0		80.7	87.0	
25 (p	o)	lare	tis							υ	G	D	actA	ပ္ပ	G rplC		G rplD	CG rplW		3G rplB	osis	
S Table 1 (continued)	Homologous gene	Mycobacterium intracellulare rpsL	Mycobacterium smegmatis LR222 rpsG	Micrococcus luteus fusA			Chlamydia trachomatis			Escherichia coli K12 fepC	Escherichia coli K12 fepG	Escherichia coli K12 fepD	Thermoanaerobacterium thermosaccharolyticum actA	Planobispora rosea ATCC 53733 rpsJ	Mycobacterium bovis BCG rplC		Mycobacterium bovis BCG rplD	Mycobacterium bovis BCG rplW		Mycobacterium bovis BCG rplB	Mycobacterium tuberculosis H37Rv Rv0705 rpsS	
35													-									
40	db Match	sp:RS12_MYCIT	sp:RS7_MYCSM	sp.EFG_MICLU			GSP: Y37841			sp:FEPC_ECOLI	sp:FEPG_ECOLI		gp:CTACTAGEN_1	sp.RS10_PLARO	sp:RL3_MYCBO		sp:RL4_MYCBO	sp.RL23_MYCBO		sp:RL2_MYCLE	sp:RS19_MYCTU	
	ORF (bp)	366	465	2115	2160	144	228	153	729	792	1035	1035	516	303	654	687	654	303	327	840	276	285
45	Terminal (nt)	523059	523533	526010	523911	526013	526894	527607	528768	528779	529592	530748	532523	533401	534090	533401	534743	535048	534746	535915	536210	535899
50	Initial (nt)	522694	523069	523896	526070	526156	527121	527759	528040	529570	530626	531782	532008	533099	533437	534087	534090	534746	535072	535076	535935	536183
	SEQ NO.	4053	4054	4055	4056	4057	4058	4059	4060	4061	4062	4063	4064	4065	4066	4067	4068	4069	4070	4071	4072	4073
55	SEQ NO.		554	555	556	257	558	559	999	561	562	563	564	565	566	567	568	569	570	571	572	573

5	.,	Function	50S ribosomal protein L22	30S ribosomal protein S3	50S ribosomal protein L16	50S ribosomal protein L29	30S ribosomal protein S17				50S ribosomal protein L14	50S ribosomal protein L24	50S ribosomal protein L5		2,5-diketo-D-gluconic acid reductase		formate dehydrogenase chain D	molybdopterin-guanine dinucleotide biosynthesis protein	formate dehydrogenase H or alpha chain			ABC transporter ATP-binding protein		
15		Matched length (a.a.)	109 50	239 30	137 50	67 50	82 30				122 50	105 50	183 50		260 2,		298 fo	94 bic	756 for			624 AE		
20		Similarity (%)	91.7	. 91.2	88.3	88.1	89.0				95.1	91.4	92.3		74.2		59.7	68.1	53.4			52.6		
		Identity (%)	74.3	77.4	69.3	65.7	69.5				83.6	76.2	73.6		52.3		28.9	37.2	24.3		_	26.9		
25	Table 1 (continued)	s gene	berculosis V	vis BCG rpsC	vis BCG rpIP	vis BCG rpmC	wis BCG rpsQ				berculosis IN	berculosis  X	s rplE		sp.		Jenes fdhD	licolor A3(2)	ſΓ			berculosis ppD		
30	Table 1 (	Homologous gene	Mycobacterium tuberculosis H37Rv Rv0706 rpIV	Mycobacterium bovis BCG rpsC	Mycobacterium bovis BCG rplP	Mycobacterium bovis BCG rpmC	Mycobacterium bovis BCG rpsQ				Mycobacterium tuberculosis H37Rv Rv0714 rplN	Mycobacterium tuberculosis H37Rv Rv0715 rplX	Micrococcus luteus rpfE		Corynebacterium sp.		Wolinella succinogenes fdhD	Streptomyces coelicolor A3(2) SCGD3.29c	Escherichia coli fdfF			Mycobacterium tuberculosis H37Rv Rv1281c oppD		
35 40		db Match	Sp.RL22_MYCTU	Sp.RS3_MYCBO N	Sp.RL16_MYCBO N	sp:RL29_MYCBO N	sp.RS17_MYCBO N				sp.RL14_MYCTU	Sp.RL24_MYCTU	SP.RL5_MICLU N		sp:2DKG_CORSP   C		SP:FDHD_WOLSU V	gp:SCGD3_29	Sp.FDHF_ECOLI E			sp:YC81_MYCTU   A		
		ORF (bp)	360	744	414	228	276	294	318	969	366	312	573	1032	807	492	915	336	2133	756	804	1662	1146	1074
45		Terminal (nt)	536576	537322	537741	537971	538252	537974	538381	538718	540106	540423	540998	542079	542090	542921	543415	544335	544757	548084	548187	548990	550699	551854
50		Initial (nt)	536217	536579	537328	537744	537977	538267	538698	539413	539741	540112	540426	541048	542896	543412	544329	544670	546889	547329	548990	550651	551844	552927
		SEQ NO.	<del></del>	4075	4076	4077	4078	4079	4080	4081	4082	4083	4084	4085	4086	4087	4088	4089	4090	4091	4092	4093	4094	4095
55		SEQ NO.	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	290	591	592	593	594	595

EP 1 108 790 A2

	_							_		-		$\overline{}$								-				
5		Function	otein	otein	protein S8	protein L6	protein L18	protein S5	protein L30	protein L15		methylmalonic acid semialdehyde dehydrogenase	H	novel two-component regulatory system	aldehyde dehydrogenase or betaine aldehyde dehydrogenase		_		sin	p-cumic alcohol dehydrogenase	otein	phosphoenolpyruvate synthetase	phosphoenolpyruvate synthetase	20
10		L.	hypothetical protein	hypothetical protein	30S ribosomal protein S8	50S ribosomal protein L6	50S ribosomal protein L18	30S ribosomal protein S5	50S ribosomal protein L30	50S ribosomal protein L15		methylmalonic a dehydrogenase		novel two-comp system	aldehyde dehydrogenase aldehyde dehydrogenase			reductase	2Fe2S ferredoxin	p-cumic alcoho	hypothetical protein	phosphoenolpy	phosphoenolpy	cytochrome P450
15		Matched length (a.a.)	405	150	132	179	110	171	55	143		128		125	487			409	107	257	90	629	378	422
20		Similarity (%)	50.4	66.7	7.76	87.7	6.06	88.3	76.4	87.4		68.8		52.0	71.5			71.6	66.4	70.8	26.0	45.0	66.7	65.2
		Identity (%)	24.7	42.7	75.8	59.2	67.3	67.8	54.6	66.4		46.9		47.0	41.7			41.1	47.7	35.8	50.0	22.9	38.6	34.8
25	ned)	ЭС	AF1398	۶۲			~ 1	111	Гц			r msdA		carR	sno			7	s fdxE	/mB	PE0029	1 DSM	1 DSM	lis thcB
30	Table 1 (continued)	Homologous gene	Archaeoglobus fulgidus AF1398	Deinococcus radiodurans DR0763	Micrococcus luteus	Micrococcus luteus	Micrococcus luteus rplR	Micrococcus luteus rpsE	Escherichia coli K12 rpmJ	Micrococcus luteus rplO		Streptomyces coelicolor msdA		Azospirillum brasilense carR	Rhodococcus rhodochrous plasmid pRTL1 orf5	-		Sphingomonas sp. redA2	Rhodobacter capsulatus fdxE	Pseudomonas putida cymB	Aeropyrum pernix K1 APE0029	Pyrococcus furiosus Vc1 DSM 3638 ppsA	Pyrococcus furiosus Vc1 DSM 3638 ppsA	Rhodococcus erythropolis thcB
35				1_13			CLU	27	ECOLI	CLU		4		3A_2	ш			8	<b>6</b>	15_2				U
40		db Match	pir.E69424	gp:AE001931_13	pir.S29885	pir.S29886	sp:RL18_MICLU	sp:RS5_MICLU	sp.RL30_EC	sp:RL15_MICLU		prf.2204281A		GP:ABCARRA_2	prf.2516398E			prf.2411257B	prf.2313248B	gp:PPU24215	PIR:H72754	pir.JC4176	pir.JC4176	prf.2104333G
		ORF (bp)	1182	468	396	534	402	633	183	444	729	321	363	456	1491	735	306	1266	318	744	213	1740	1080	1290
45		Terminal (nt)	552948	554452	555726	556282	556690	557366	557555	558008	556860	558197	558607	560260	559144	560634	562937	561368	562646	562993	564083	563732	565680	566799
50		Initial (nt)	554129	554919	555331	555749	556289	556734	557373	557565	557588	558517	558969	559805	560634	561368	562632	562633	562963	563736	563871	565471	566759	568088
		SEQ NO. (a.a.)	4096	4097	4098	4099	4100	4101	4102	4103	4104	4105	4106	4107	4108	4109	4110	4111	4112	4113	4114	4115	4116	4117
		SEQ NO.	969	597	598	599	900	601	602	603	604	605	909	607	809	609	610	611	612	613	614	615	616	617

							_											
5	<i>t</i> -		uo	ssor			ptidase		factor IF-1	in S13	in S11	iin S4	oha subunit		in L17	thase A	ane protein	
10			Function	transcriptional repressor	adenylate kinase		methionine aminopeptidase		translation initiation factor IF-1	30S ribosomal protein S13	30S ribosomal protein S11	30S ribosomal protein S4	RNA polymerase alpha subunit		50S ribosomal protein L17	pseudouridylate synthase A	hypothetical membrane protein	
15			Matched length (a.a.)	256	184		253		72	122	134	132	311		122	265	786	
20			Similarity (%)	0.99	81.0		74.7		96.0	91.0	93.3	93.9	77.8		77.1	61.1	51.2	
			Identity (%)	28.5	48.9		43.1		77.0	66.4	81.3	82.6	51.1		51.6	37.0	24.8	
25		ntinued)	gene	arotovora	adk		map			us HB8	olor A3(2)	erculosis sD	rpoA		2 rpIQ	2 truA	erculosis	
30		Table 1 (continued)	Homologous gene	Erwinia carotovora carotovora kdgR	Micrococcus luteus adk		Bacillus subtilis 168 map		Bacillus subtilis infA	Thermus thermophilus HBB rps13	Streptomyces coelicolor A3(2) SC6G4.06. rpsK	Mycobacterium tuberculosis H37Rv RV3458C rpsD	Bacillus subtilis 168 rpoA		Escherichia coli K12 rplQ	Escherichia coli K12 truA	Mycobacterium tuberculosis H37Rv Rv3779	
<i>35</i> <i>40</i>			db Match	prf.2512309A k	sp:KAD_MICLU		SP. AMPM_BACSU E		pir.F69644	prf;2505353B	sp.RS11_STRCO	prf.2211287F	Sp.RPOA_BACSU		sp:RL17_ECOLI	sp:TRUA_ECOLI	pir.G70695	
			ORF (bp)	804	543	612	792	828	216	366	402	603	1014	156	489	867	2397.	456
45			Terminal (nt)	568272	571316	570756	572267	573176	573622	574181	574588	575217	576351	575211	576898	577923	580429	580436
50			Initial (nt)	569075	570774	571367	571476	572349	573407	573816	574187	574615	575338	575366	576410	577057	578033	500001
			SEQ NO. (a.a.)	4118	4119	4120	4121	4122	4123	4124	4125	4126	4127	4128	4129	4130	4131	1133
55			SEQ NO.	618	619	620	621	622	623	624	625	626	627	628	629	630	631	623

cyclopropane-fatty-acyl-phospholipid synthase hypothetical membrane protein cell elongation protein hypothetical protein 505 423 485 5 53.8 50.9 56.0 59.0 28.0 27.4 22.8 30.7 Streptomyces coelicolor A3(2) SCL2.30c Arabidopsis thaliana CV DIM Mycobacterium tuberculosis H37Rv Rv0283 Escherichia coli K12 cfa Sp:DIM\_ARATH 1353 Sp:CFA\_ECOLI gp:SCL2\_30 pir.A70836 1545 1257 426 632 | 4132 | 580891 | 580436 | 456 303 580919 585620 586248 584228 582662 581221 585823 581406 582684 584268 4133 4134 4135 4136 4137 633 635 636 637

	0	C		- 4	-1	- 1	-	- 4								- 1		1	1	C 2
5			oteinase	e protein	e protein					target ESAT	L13	S9	nutase							
10		Function	high-alkaline serine proteinase	hypothetical membrane protein	hypothetical membrane protein				hypothetical protein	early secretory antigen target ESAT- 6 protein	50S ribosomal protein L13	30S ribosomal protein S9	phosphoglucosamine mutase		hypothetical protein			hypothetical protein	alanine racemase	hypothetical protein
15		Matched tength (a.a.)	273	516	1260				103	80	145	181	450		318			259	368	154
20		Similarity (%)	58.0	50.6	38.4				69.6	81.3	82.1	72.4	76.4		45.6			72.2	68.5	78.6
		Identity (%)	31.3	24.0	65.0				31.1	36.3	58.6	49.2	48.9		29.3			44.0	41.6	48.7
25	Table 1 (continued)	s gene	IS	icolor A3(2)	erculosis				erculosis	erculosis	icolor A3(2)	icolor A3(2)	ıreus		PCC6803			ırae	oerculosis alr	erculosis
30	Table 1 (c	Homologous gene	Bacillus alcalophilus	Streptomyces coelicolor A3(2) SC3C3.21	Mycobacterium tuberculosis H37Rv Rv3447c				Mycobacterium tuberculosis H37Rv Rv3445c	Mycobacterium tuberculosis	Streptomyces coelicolor A3(2) SC6G4.12. rpIM	Streptomyces coelicolor A3(2) SC6G4.13. rpsl	Staphylococcus aureus femR315		Synechocystis sp. PCC6803 slr1753			Mycobacterium leprae B229_F1_20	Mycobacterium tuberculosis H37Rv RV3423C alr	Mycobacterium tuberculosis H37Rv Rv3422c
35			Ba	Scr	₹ 13			_	₹ E	Σ	SC St	\$ S	fe St		S <sub>l</sub>		_	83	ΣÏ	ΣÏ
40		db Match	SP.ELYA_BACAO	pir.T10930	pir.E70977				pir.C70977	prf.2111376A	sp:RL13_STRCO	sp:RS9_STRCO	prf.2320260A		pir.S75138	-		pir.S73000	sp:ALR_MYCTU	sp:Y097_MYCTU
		ORF (bp)	1359	1371	3567	822	663	906	324	288	441	546	1341	303	1509	573	234	855	1083	495
45		Terminal (nt)	586399	587645	592862	589590	589898	593761	594258	594580	595379	595927	597449	598194	599702	598778	599932	600022	602053	602574
50		Initial (nt)	587757	589015	589296	590411	590560	592862	593935	594293	594939	595382	596109	597892	598194	599350	599699	600876	600971	602080
		SEQ NO.	4138	4139	4140	4141	4142	4143	4144	4145	4146	4147	4148	4149	4150	4151	4152	4153	4154	4155
55		SEQ NO.	+	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655

EP 1 108 790 A2

	_																			
5			protein			ne N-	lopeptidase				S					a factor				
10		Function	hypothetical membrane protein	proline iminopeptidase	hypothetical protein	ribosomal-protein-alanine N- acetytransferase	O-sialoglycoprotein endopeptidase	hypothetical protein			heat shock protein groES	heat shock protein groEL	hypothetical protein	hypothetical protein	regulatory protein	RNA polymerase sigma factor		hypothetical protein	IMP dehydrogenase	hypothetical protein
15		Matched length (a.a.)	550	411	207	132	319	571			100	537	9/	138	94	174		116	504	146
20		Similarity (%)	66.2	77.6	75.4	59.9	75.2	59.4			94.0	85.1	56.0	45.0	88.3	81.6		8.69	93.9	53.0
		Identity (%)	28.9	51.3	52.2	30.3	46.1	38.4		<i>:</i>	76.0	63.3	50.0	34.0	64.9	55.2		41.4	80.8	39.0
25	Table 1 (continued)	ns gene	12 yidE	ı shermanii pip	berculosis	12 riml	olytica cp	berculosis			berculosis mopB	prae 5E1	berculosis	berculosis	negmatis	iberculosis sigD		prae	TCC 6872	oshii PH0308
30	Table 1 (	Homologous gene	Escherichia coli K12 yidE	Propionibacterium shermanii	Mycobacterium tuberculosis H37Rv Rv3421c	Escherichia coli K12 riml	Pasteurella haemolytica SEROTYPE A1 gcp	Mycobacterium tuberculosis H37Rv Rv3433c			Mycobacterium tuberculosis H37Rv RV3418C mopB	Mycobacterium leprae B229_C3_248 groE1	Mycobacterium tuberculosis	Mycobacterium tuberculosis	Mycobacterium smegmatis whiB3	Mycobacterium tuberculosis H37Rv Rv3414c sigD		Mycobacterium leprae B1620_F3_131	Corynebacterium ammoniagenes ATCC 6872 guaB	Pyrococcus horikoshii PH0308
35			ш	Δ.		ш								€.	23				Oes	<u>-</u>
40		db Match	sp:YIDE_ECOLI	gp:PSJ00161_1	sp:Y098_MYCTU	sp.RIMI_ECOL!	sp.GCP_PASHA	sp:Y115_MYCTU			sp:CH10_MYCTU	sp:CH61_MYCLE	GP.MSGTCWPA_1	GP:MSGTCWPA_	gp:AF073300_1	sp:Y09F_MYCTU		sp:Y09H_MYCLE	gp:AB003154_1	PIR:F71456
		ORF (bp)	1599	1239	675	507	1032	1722	429	453	297	1614	255	1158	297	564	1026	378	1518	627
45		Terminal (nt)	604409	605708	606392	606898	607936	609679	610175	609816	610644	612272	610946	611109	612418	613719	614747	614803	616853	615605
50		Initial (nt)	602811	604470	605718	606392	606905	607958	609747	610268	610348	610659	611200	612266	612714	613156	613722	615180	615336	616231
		SEQ NO.	4156	4157	4158	4159	4160	4161	4162	4163	4164	4165	4166	4167	4168	4169	4170	4171	4172	4173
55		SEQ NO.		657	658	629	099	661	662	663	664	665	999	299	999	699	670	671	672	673

EP 1 108 790 A2

5		Function	IMP dehydrogenase	hypothetical membrane protein	glutamate synthetase positive regulator	GMP synthetase				hypothetical membrane protein	two-component system sensor histidine kinase	transcriptional regulator or extracellular proteinase response regulator	-			hypothetical protein	hypothetical protein	-	hypothetical protein	hypothetical membrane protein	
15		Matched length (a.a.)	381	274	262	517				513	411	218				201	563		275	288	
20		Similarity (%)	86.1	67.5	58.4	92.8				39.6	48.7	65.1				64.2	64.1		62.9	58.3	
		Identity (%)	70.9	38.0	29.0	81.6				20.5	26.8	33.5				30.9	37.5		33.8	27.8	
25	Table 1 (continued)	us gene	TCC 6872	12 ybiF	၁	аА				licolor A3(2)	licolor A3(2)	i8 degU				berculosis	berculosis		licolor A3(2)	odurans	
30	Table 1 (	Homologous gene	Corynebacterium ammoniagenes ATCC 6872	Escherichia coli K12 ybiF	Bacillus subtilis gltC	Corynebacterium ammoniagenes guaA				Streptomyces coelicolor A3(2)	Streptomyces coelicolor A3(2) SC6E10.15c	Bacillus subtilis 168 degU				Mycobacterium tuberculosis H37Rv Rv3395c	Mycobacterium tuberculosis H37Rv Rv3394c		Streptomyces coelicolor A3(2) SC5B8.20c	Deinococcus radiodurans DR0809	
35 40		db Match	gp:AB003154_2 a	sp:YBIF_ECOLI E		sp:GUAA_CORAM				gp:SCD63_22 S	-15	sp.DEGU_BACSU   E				pir.B70975	Pir.A70975		gp:SC5B8_20	gp:AE001935_7	
40		ORF (bp)	1122 gp:A	921 sp:Y	909 prf.1	1569 sp.G	663	441	189	1176 gp:S	1140 gp:S	G:08 069	324	489	963	825 pir.B	1590 pir.A	099	861 gp:S	861 gp:A	390
45		Terminal O (tr)	618094 11	618093 9	619994 9	621572 16	620264 6	622157 4	622457 1	622460 11	624939 11	625674 6	626000 3	626070 4	626577 9	628551 8	630140 18	630151 6	631809 8	631824 8	632690 3
50		Initial (nt)	616973	619013	619086	620004	620926	621717	62229	623635	623800	624985	625677	626558	627539	627727	628551	630810	630949	632684	633079
		SEQ NO.	4174	4175	4176	4177	4178	4179	4180	4181	4182	4183	4184	4185	4186	4187	4188	4189	4190	4191	4192
55		SEQ NO.	674	675	929	677	678	679	989	681	682	683	684	685	989	687	688	689	069	691	692

5			Function	hypothetical membrane protein	aturase	hase	transmembrane transport protein	geranylgeranyl pyrophosphate (GGPP) synthase	transcriptional regulator (MarR family)	outer membrane lipoprotein	ırotein	es	ferase	ter	ter		ter		rter		rase III	protein
10				hypothetical n	phytoene desaturase	phytoene synthase	transmembra	geranylgeranyl py (GGPP) synthase	transcriptiona family)	outer membra	hypothetical protein	DNA photolyase	glycosyl transferase	ABC transporter	ABC transporter		ABC transporter		ABC transporter	lipoprotein	DNA polymerase III	hypothetical protein
15		Matched	length (a.a.)	95	524	288	722	367	188	145	462	497	205	897	223		206		346	268	1101	159
20		واسان	Similarity (%)	67.4	76.2	71.2	75.6	63.8	68.1	62.1	74.2	63.2	53.7	54.9	72.2		75.2		75.4	67.2	57.5	62.3
		1	(%)	36.8	50.4	42.0	48.6	32.7	38.3	33.1	48.7	40.0	25.9	24.3	35.4		35.9		43.6	28.7	30.2	41.5
25	nued)		:ue	E	ATCC	ATCC	or A3(2)	erE		OS60 blc		АТСС	\$1K	or A3(2)	Q		e l		0 abc	ae 3e	ıaE	or A3(2)
30	Table 1 (continued)		Hamologous gene	Mycobacterium marinum	Brevibacterium linens ATCC 9175 crtl	Brevibacterium linens ATCC 9175 crtB	Streptomyces coelicolor A3(2) SCF43A.29c	Brevibacterium linens crtE	Brevibacterium linens	Citrobacter freundii blc OS60 blc	Brevibacterium linens	Brevibacterium linens ATCC 9175 cpd1	Streptococcus suis cps1K	Streptomyces coelicolor A3(2) SCE25.30	Bacillus subtilis 168 yvrO		Helicobacter pylori abcD		Escherichia coli TAP90 abc	Haemophilus influenzae SEROTYPE B hlpA	Thermus aquaticus dnaE	Streptomyces coelicolor A3(2) SCE126.11
35 40			db Match	ap:MMU92075 3		gp:AF139916_2	gp:SCF43A_29	gp:AF139916_11	gp:AF139916_14	Sp.BLC CITFR	gp:AF139916_1	gp:AF139916_5	gp:AF155804_7	gp:SCE25_30	prf:2420410P		prf:2320284D		sp:ABC_ECOLI	sp.HLPA_HAEIN	prf:2517386A	gp:SCE126_11
			ORF (bp)	396	<del></del>	912	2190	1146	585	648	1425	1404	753	2415	717	153	999	846	1080	897	3012	447
45			Terminal (nt)	633079	633532	635178	636089	638317	640208	640232	642557	642556	644778	645176	647593	648315	648440	650187	649114	650392	654612	655122
50			Initial (nt)	633474	635175	636089	638278	639462	639624	640879	641133	643959	644026	647590	648309	648467	649105	649342	650193	651288	651601	654676
		0.0	NO SEC	4103	4194	4195	4196	4197	4198	4199	4200	4201	4202	4203	4204	4205	4206	4207	4208	4209	4210	4211
55			NO S	502	694	695	969	697	869	909	8 8	707	702	703	704	705	706	707	708	709	710	711

EP 1 108 790 A2

																		<del></del>			
5		Function	hypothetical membrane protein		transcriptional repressor	hypothetical protein		transcriptional regulator (Sir2 family)	hypothetical protein	iron-regulated lipoprotein precursor	rRNA methylase	methylenetetrahydrofolate dehydrogenase	hypothetical membrane protein	hypothetical protein		homoserine O-acetyltransferase	O-acetylhomoserine sulfhydrylase	carbon starvation protein		hypothetical protein	
٠		70	hypc		tran	hyp		tran	hyp	<u>ro</u>	Ϋ́ Ng	deh deh	hyp	hyp	-	hon	0.9	cari		d Y	4
15		Matched length (a.a.)	468		203	264		245	157	357	151	278	80	489		379	429	069		8	
20		Similarity (%)	56.0		76.4	61.7		71.8	78.3	62.2	86.1	87.4	76.3	63.2		99.5	76.2	78.4		66.0	
		Identity (%)	26.1		50.3	34.9		42.5	45.2	31.1	65.9	70.9	31.3	34.0		99.5	49.7	53.9		40.0	
25	itinued)	gene	olor A3(2)		culosis	olor A3(2)		lus AF1676	olor A3(2)	htheriae	rculosis J	rculosis )	le	olor A3(2)		ıtamicum	etY	cstA		XijX	
30	Table 1 (continued)	Homologous gene	Streptomyces coelicolor A3(2) SCE9.01		Mycobacterium tuberculosis H37Rv Rv2788 sirR	Streptomyces coelicolor A3(2) SCG8A.05c		Archaeoglobus fulgidus AF1676	Streptomyces coelicolor A3(2) SC5H1.34	Corynebacterium diphtheriae irp1	Mycobacterium tuberculosis H37Rv Rv3366 spoU	Mycobacterium tuberculosis H37Rv Rv3356c folD	Mycobacterium leprae MLCB1779.16c	Streptomyces coelicolor A3(2) SC66T3.18c		Corynebacterium glutamicum metA	Leptospira meyeri metY	Escherichia coli K12 cstA		Escherichia coli K12 yjiX	
35			Str		ΣÏ	38		Ā	क छ		ΣÏ	ΣÏ	≨ <b>∑</b> ∞.	क्र छ			l e				
40		db Match	gp:SCE9_1		pir.C70884	gp:SCG8A_5		pir:C69459	gp:SC5H1_34	gp:CDU02617_1	pir.E70971	pir:C70970	gp:MLCB1779_	gp:SC66T3_18		gp:AF052652_1	prf:2317335A	sp.CSTA_ECOLI		sp:YJIX_ECOLI	
		ORF (bp)	1413	738	699	798	138	774	492	966	471	852	255	1380	963	1131	1311	2202	609	201	609
45		Terminal (nt)	656534	655097	657215	657205	658142	658928	659424	660538	660650	662017	662374	662382	664126	665183	666460	670465	669445	670672	671045
50		Initial (nt)	655122	655834	656547	658002	658005	658155	658933	659543	661120	661166	662120	663761	665088	666313	667770	668264	670053	670472	671653
		SEO NO (a.a.)	4212	4213	4214	4215	4216	4217	4218	4219	4220	4221	4222	4223	4224	4225	4226	4227	4228	4229	4230
55		SEQ NO. DNA)	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730

hypothetical protein

219

82.8

56.2

Haemophilus influenzae Rd H11240

Sp: YICG\_HAEIN

657 195

688335 688007

748 | 4248 | 688141 747 | 4247 | 687351

	- 1																		_
5		Function	hypothetical protein	carboxy phosphoenolpyruvate mutase	citrate synthase		hypothetical protein	•	L-malate dehydrogenase	regulatory protein		vibriobactin utilization protein	ABC transporter ATP-binding protein	ABC transporter	ABC transporter	iron-regulated lipoprotein precursor	chloramphenicol resistance protein	catabolite repression control protein	
15		Matched length (a.a.)	317	281	380		53		338	226		284	269	339	330	356	395	303	
20		Similarity (%)	86.4	76.2	81.3		62.3		67.5	62.8		54.2	1.28	86.4	88.2	82.3	9.69	58.1	
		Identity (%)	71.0	41.6	56.1		34.0		37.6	26.1		25.4	55.4	56.3	63.0	53.1	32.2	30.4	
25 :	Table 1 (continued)	ss gene	perculosis	roscopicus	negmatis		12 yneC		ervidus V24S	stearothermophilus T-6		SAWA 395	diphtheriae	diphtheriae	diphtheriae	diphtheriae	szuelae cmlv	uginosa crc	
	Table 1 (	Homologous gene	Mycobacterium tuberculosis H37Rv Rv1130	Streptomyces hygroscopicus	Mycobacterium smegmatis ATCC 607 gltA		Escherichia coli K12 yneC		Methanothermus fervidus V24S mdh	Bacillus stearother		Vibrio cholerae OGAWA 395 viuB	Corynebacterium diphtheriae irp1D	Corynebacterium diphtheriae irp1C	Corynebacterium diphtheriae irp1B	Corynebacterium diphtheriae irp1	Streptomyces venezuelae cmlv	Pseudomonas aeruginosa crc	1
35 40		db Match	pir.C70539	prf. 1902224A	sp.CISY_MYCSM		Sp:YNEC_ECOLI E		Sp:MDH_METFE	prf:2514353L		sp:VIUB_VIBCH	gp:AF176902_3	gp:AF176902_2	gp:AF176902_1	gp:CDU02617_1		prf.2222220B	
		ORF (bp)	954	912	1149	930	192	672	1041	720	702	897	807	1059	966	1050	1272	912	
45		Terminal (nt)	672653	673576	674756	672710	674799	675846	675082	676218	677047	680131	681040	681846	682871	683876	686380	687346	
50		Initial (nt)	671700	672665	673608	673639	674990	675175	676122	676937	677748	681027	681846	682904	683866	684925	685109	686435	
		SEQ NO.	4231	4232	4233	4234	4235	4236	4237	4238	4239	4240	4241	4242	4243	4244	4245	4246	
55		SEQ NO. (DNA)	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	

5		Function		ferrichrome ABC transporter	hemin permease	tryptophanyl-tRNA synthetase	hypothetical protein		penicillin-binding protein ob precursor	hypothetical protein	hypothetical protein			uracii phosphoribosyltransterase	bacterial regulatory protein, laci family	N-acyl-L-amino acid amidohydrolase or peptidase	phosphomannomutase	dihydrolipoamide dehydrogenase	pyruvate carboxylase	hypothetical protein	hypothetical protein
15		Matched length (a.a.)		244	346	331	278		301	417	323			509	77	385	561	468	1140	263	127
20		Similarity (%)		73.8	69.1	79.8	72.3		57.5	70.7	52.6			72.3	66.2	80.5	53.8	65.0	100.0	60.1	6.99
		Identity (%)		45.1	38.7	54.4	37.1		30.9	34.1	29.4			46.4	41.6	51.4	22.1	31.6	100.0	26.2	30.7
30	Table 1 (continued)	Homologous gene		Corynebacterium diphtheriae hmuV	Yersinia enterocolitica hemU	Escherichia coli K12 trpS	Escherichia coli K12 yhjD		Salmonella typhimurium LT2 dacD	Mycobacterium tuberculosis H37Rv Rv3311	Streptomyces coelicolor A3(2) SC6G10.08c			Lactococcus tactis upp	Streptomyces coelicolor A3(2) SC1A2.11	Mycobacterium tuberculosis H37Rv Rv3305c amiA	Mycoplasma pirum BER manB	Halobacterium volcanii ATCC 29605 lpd	Corynebacterium glutamicum strain21253 pyc	Mycobacterium tuberculosis H37Rv Rv1324	Streptomyces coelicalor A3(2) SCF11.30
35				Coryne	Yersi	Esch	Esch		Salmo dacD	Myco H37F	Strep			Lacte	Strep SC1/	Mycc H37F	T		Cory		Stre
40		db Match		gp:AF109162_3	pir.S54438	sp:SWW_ECOLI	sp:YHJD_ECOLI		sp.DACD_SALTY	pir.F70842	gp:SC6G10_8			Sp:UPP_LACLA	gp:SC1A2_11	pir.H70841	Sp.MANB_MYCPI		prf.2415454A	sp:YD24_MYCTU	gp:SCF11_30
		ORF (bp)	975	780	1017	1035	1083	903	1137	1227	858	195	351	633	384	1182	1725	1407	3420	870	486
45		Terminal (nt)	688916	689917	690706	692916	694110	695074	695077	696769	698065	699266	698922	699913	700381	703262	700384	704811	708630	709708	710278
50		Initial (nt)	689890	969069	691722	691882	693028	694172	696213	697995	698922	699072	699272	699281	866669	702081	702108		705211	708839	709793
		SEQ NO.	4249	4250	4251	4252	4253	4254	4255	4256	4257	4258	4259	4260	4261	4262	4263	4264	4265	4266	4267
55		SEQ NO DNA)	+	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	792

5 10		Function	hypothetical protein	thioredoxin reductase	PrpD protein for propionate catabolism	carboxy phosphoenolpyruvate mutase	hypothetical protein	citrate synthase		hypothetical protein			thiosulfate sulfurtransferase	hypothetical protein	hypothetical protein	hypothetical membrane protein	hypothetical protein	hypothetical protein	detergent sensitivity rescuer or carboxyl transferase	detergent sensitivity rescuer or carboxyl transferase
15		Matched length (a.a.)	381	305	521	278	96	383		456			225	352	133	718	192	63	537	543
20		Similarity (%)	69.0	59.3	49.5	74.5	47.0	78.9		72.6			100.0	79.8	76.7	63.4	66.2	69.8	100.0	100.0
		Identity (%)	44.6	24.6	24.0	42.5	39.0	54.6		40.8			100.0	61.1	51.1	35.1	31.8	33.3	8.66	9.66
<i>25</i> <i>30</i>	Table 1 (continued)	Homologous gene	Bacillus subtilis 168 yciC	Bacillus subtilis IS58 trxB	Salmonella typhimurium LT2 prpD	Streptomyces hygroscopicus	Aeropyrum pernix K1 APE0223	Mycobacterium smegmatis ATCC 607 gltA		Mycobacterium tuberculosis H37Rv Rv1129c			Corynebacterium glutamicum ATCC 13032 thtR	Campylobacter jejuni Cj0069	Mycobacterium leprae MLCB4.27c	Mycobacterium tuberculosis H37Rv Rv1565c	Escherichia coli K12 yceF	Mycobacterium leprae B1308- C3-211	Corynebacterium glutamicum AJ11060 dtsR2	Corynebacterium glutamicum AJ11060 dtsR1
35		db Match	pir.B69760 Ba	sp.TRXB_BACSU B	SP:PRPD_SALTY Br	prf:1902224A St	PIR:E72779 A	Sp.CISY_MYCSM A.		pir.B70539 H			Sp.THTR_CORGL A.	gp:CJ11168X1_62 C	gp:MLCB4_16 M	pir.G70539 H	Sp:YCEF_ECOLI E	prf.2323363CF C:	gp:AB018531_2 A.	pir.JC4991 A.
		ORF (bp)	1086	924	1494	888	378	1.182	375	1323	246	1359	903	1065	414	2148	591	246	1611	1629
45		Terminal (nt)	710520	712647	714231	715145	714380	716283	716286	716687	718350	720016	720547	722841	722925	725559	725872	726470	726742	728696
50		· Initial (nt)	711605	711724	712738	714258	714757	715102	716660	718009	718105	718658	721449	721777	723338	723412	726462	726715	728352	730324
		SEQ NO.	4268	4269	4270	4271	4272	4273	4274	4275	4276	4277	4278	4279	4280	4281	4282	4283	4284	4285
55		SEQ NO. (DNA)	768	769	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785

	_					<del></del>						-					$\neg$		$\neg$
5		Function	bifunctional protein (biotin synthesis repressor and biotin acetyl-CoA carboxylase ligase)	hypothetical membrane protein	5'-phosphoribosyl-5-amino-4- imidasol carboxylase	K+-uptake protein			5'-phosphoribosyl-5-amino-4- imidasol carboxylase	hypothetical protein	hypothetical protein	nitrilotriacetate monooxygenase	transposase (ISA0963-5)	glucose 1-dehydrogenase	hypothetical membrane protein		hypothetical protein	hypothetical protein	
			bifuncti repress carbox	hypoth	5'-phos imidas	K+-upt	-		5'-phos imidas	hypoth	$\overline{}$				hypoth	$\dashv$	寸		
15		Matched length (a.a.)	293	165	394	628			147	152	255	426	303	256	96		175	142	
20		Similarity (%)	61.8	58.8	83.8	73.6			93.2	60.5	70.6	73.0	52.5	64.8	68.8		66.3	76.8	
		Identity (%)	28.7	23.0	0.69	41.1			85.7	36.2	42.8	43.2	23.4	31.3	29.2		28.6	35.9	
25	ned)	o.	4	osis	872	0			1872	Ę	r A3(2)	202		M 1030	1SB8		<u>B</u>	r A3(2)	
30	Table 1 (continued)	Homologous gene	Escherichia coli K12 birA	Mycobacterium tuberculosis H37Rv Rv3278c	Corynebacterium ammoniagenes ATCC 6872 purK	Escherichia coli K12 kup			Corynebacterium ammoniagenes ATCC 6872 purE	Actinosynnema pretiosum	Streptomyces coelicolor A3(2) SCF43A.36	Chelatobacter heintzii ATCC 29600 ntaA	Archaeoglobus fulgidus	Bacillus megaterium IAM 1030 gdhll	Thermotoga maritima MSB8 TM1408		Bacillus subtilis 168 ywjB	Streptomyces coelicolor A3(2) SCJ9A.21	
35 40		db Match	sp.BIRA_ECOLI	pir.G70979	sp:PURK_CORAM	Sp:KUP_ECOLI			sp:PUR6_CORAM	gp:APU33059_5	gp:SCF43A_36	sp:NTAA_CHEHE	pir.A69426	sp:DHG2_BACME	pir.A72258		sp:YWJB_BACSU	gp:SCJ9A_21	
		ORF (bp)	864	486	1161	1872	615	357	495	453	792	1314	1500	789	369	342	267	420	222
45	,	Terminal (nt)	731299	731797	733017	734943	733183	735340	735896	736351	737204	737216	738673	740228	741765	742195	741818	742828	742831
50		Initial (nt)	730436	731312	731857	733072	733797	734984	735402	735899	736413	738529	740172		741397	741854	742384		743052
		SEQ.	(3.3.)	4287	4288	4289	4290	4291	4292	4293	4294	4295	4296	4297	4298	4299	4300	4301	4302
55		SEQ	786	787	788	789	290	791	792	793	794	795	796	797	798	799	800	801	802

5
10
15
<b>20</b>
25
30
35
40
45
50

_		<del></del> -		_		_		$\overline{}$		$\neg \tau$							$\neg$	$\neg \neg$				
	Function	trehalose/maltose-binding protein	trehalose/maltose-binding protein		trehalose/maltose-binding protein		ABC transporter ATP-binding protein (ABC-type sugar transport protein) or cellobiose/maltose transport protein		RNA helicase			hypothetical protein	hypothetical protein	DNA helicase II					RNA helicase	hypothetical protein	RNA polymerase associated protein (ATP-dependent helicase)	
	Matched length (a.a.)	271	306		417		332		1783			240	720	701					2033	869	873	
	Similarity (%)	75.3	70.3		62.4		73.9		49.9			59.2	62.5	41.1					45.8	53.2	48.6	
	Identity (%)	42.4	37.3		30.9		57.2		25.1			31.7	30.0	20.7					22.4	24.4	23.1	
Table 1 (continued)	Homologous gene	Thermococcus litoralis malG	Thermococcus litoralis malF		Thermococcus litoralis malE		Streptomyces reticuli msiK		Deinococcus radiodurans R1 DRB0135			Mycobacterium tuberculosis H37Rv Rv3268	Helicobacter pylori J99 jhp0462	Escherichia coli K12 uvrD					Streptomyces coelicolor SCH5.13	Halobacterium sp. NRC-1 plasmid pNRC100 H1130	Escherichia coli K12 hepA	
	db Match	prf.2406355C	prf.2406355B		prf.2406355A		prf.2308356A		pir:B75633			pir.E70978	pir.C71929	sp.UVRD_ECOLI					pir.T36671	pir.T08313	sp:HEPA_ECOLI	
	ORF (bp)	834		468	1272	423	966	369	4800	372	3699	633	2433	1563	357	393	396	825	6207	4596	2886	
	Terminal (nt)	743067	743900	745046	745622	748442	747031	748814	748886	757434	753697	757630	758364	760906	762853	763122	762582	767367	763237	769547	774150	
	Initial (nt)	743900	744931	745513	746893	748020	748026	748446	753685	757063	757395	758262	760796	762468	762497	762730	762977	768191	769443	774142	777035	,
	SEQ	4303		<del></del>	4306	4307	4308	4309	4310	4311	4312	4313	4314	4315	4316	4317	4318	4319	4320	4321	4322	]
	SEQ		<del></del>	+	1	1	1	809		811	812	813	814	815	816	817	818	819	820	821	822	

EP 1 108 790 A2

											-						<del></del>		
5		Function	hypothetical protein	dTDP-Rha:a-D-GlcNAc- diphosphoryl polyprenol, a-3-L- rhamnosyl transferase	mannose-1-phosphate guanylyltransferase	regulatory protein	hypothetical protein	hypothetical protein	phosphomannomutase	hypothetical protein	mannose-6-phosphate isomerase			pheromone-responsive protein		S-adenosyl-L-homocysteine hydrolase			thymidylate kinase
		- œ	hypothe	dTDP-R diphosp rhamnos	mannos guanyly	regulato	hypothe	hypothe		hypothe	mannos			pherom					
15		Matched length (a.a.)	527	289	353	94	139	136	460	327	420			180		476			209
20		Similarity (%)	71.4	77.9	6.99	81.9	74.8	71.3	66.3	56.3	66.2			57.8		83.0			26.0
		Identity (%)	45.5	56.4	29.8	73.4	48.9	51.5	38.0	31.2	36.9			35.6		29.0			25.8
25	ntinued)	gene	rcutosis	gmatis	evisiae	gmatis	erculosis	color A3(2)	deo M40	erculosis	2 manA			ils plasmid		alis WAA38			idus VC-16
30	Table 1 (continued)	Homologous gene	Mycobacterium tubercutosis H37Rv Rv3267	Mycobacterium smegmatis mc2155 wbbL	Saccharomyces cerevisiae YDL055C MPG1	Mycobacterium smegmatis whmD	Mycobacterium tuberculosis H37Rv Rv3259	Streptomyces coelicolor A3(2) SCE34.11c	Satmonella montevideo M40 manB	Mycobacterium tuberculosis H37Rv Rv3256c	Escherichia coli K12 manA	-		Enterococcus faecalis plasmid pCF10 prgC		Trichomonas vaginalis WAA38			Archaeoglobus fulgidus VC-16 AF0061
<i>35</i>		db Match	pir:D70978	gp:AF187550_1	sp:MPG1_YEAST	gp:AF164439_1	pir:B70847	gp:SCE34_11	sp:MANB_SALMO	pir.B70594	sp:MANA_ECOLI			prf.1804279K		sp:SAHH_TRIVA			sp.KTHY_ARCFU
		ORF (bp)	1554	897	1044	408	456	390	1374	1005	1182	150	360	564	351	1422	708	720	609
45		Terminal (nt)	777158	779910	781171	781875	782162	783101	784557	785639	786824	787045	787983	787170	788546	790093	788719	789002	790704
50		Initial (nt)	778711	779014	780128	781468	782617	782712	783184	784635	785643	786896	787624	787733	788196	788672	789426	789721	790096
		SEQ NO.	4323	4324	4325	4326	4327	4328	4329	4330	4331	4332	4333	4334	4335	4336	4337	4338	4339
55		SEQ NO.	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839

5		Function	two-component system response regulator		two-component system sensor histidine kinase	lipoprotein	hypothetical protein		30S ribosomal protein or chloroplast precursor	preprotein translocase SecA subunit		hypothetical protein
15		Matched length (a.a.)	224 tw		484 tw	595 lip	213 hy		203 30	845 pr		170 hy
20		Similarity (%)	9.06		6.82	65.6	72.8		61.6	9.66		78.8
		Identity (%)	73.7		53.1	29.6	38.0		34.5	99.1		47.1
25	ontinued)	gene	erculosis rA		erculosis rB	erculosis 1B	erculosis		3V rps22	um jlutamicum)		erculosis
30	Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv3246c mtrA		Mycobacterium tuberculosis H37Rv Rv3245c mtrB	Mycobacterium tuberculosis H37Rv Rv3244c lpqB	Mycobacterium tuberculosis H37Rv Rv3242c		Spinacia oleracea CV rps22	Brevibacterium flavum (Corynebacterium glutamicum) MJ-233 secA		Mycobacterium tuberculosis H37Rv Rv3231c
<b>35</b>			ΣI		21	21	21			B 5.2		21
40		db Match	prf.2214304A		prf:2214304B	pir.F70592	pir.D70592		sp.RR30_SPIOL	gsp:R74093		pir.A70591
		ORF (bp)	678	684	1497	1704	588	156	663	2535	672	504
45		Terminal (nt)	791409	790738	793008	794711	795301	795292	796110	798784	799691	800200
50		Initial (nt)	790732	791421	791512	793008	794714	795447	795448	796250	799020	799697
		SEQ NO.		4341	4342	4343	4344	4345	4346	4347	4348	4349
		0 0.€	0	-	2	3	4	5	9	7	80	0

Function	two-component system response regulator		two-component system sensor histidine kinase	lipoprotein	hypothetical protein		30S ribosomal protein or chloroplast precursor	preprotein translocase SecA subunit		hypothetical protein	hypothetical protein	5-enolpyruvylshikimate 3-phosphate synthase	hypothetical protein	5-enolpyruvylshikimate 3-phosphate synthase	hypothetical protein	RNA polymerase sigma factor
Matched length (a.a.)	224		484	595	213		203	845		170	322	461	180	23	380	188
Similarity (%)	90.6		78.9	65.6	72.8		61.6	9.66		78.8	82.9	99.0	63.9	100.0	42.4	87.2
Identity (%)	73.7		53.1	29.6	38.0		34.5	99.1		47.1	64.6	0.86	38.3	100.0	21.6	61.2
Homologous gene	Mycobacterium tuberculosis H37Rv Rv3246c mtrA		Mycobacterium tuberculosis H37Rv Rv3245c mtrB	Mycobacterium tuberculosis H37Rv Rv3244c lpqB	Mycobacterium tuberculosis H37Rv Rv3242c		Spinacia oleracea CV rps22	Brevibacterium flavum (Corynebacterium glutamicum) MJ-233 secA		Mycobacterium tuberculosis H37Rv Rv3231c	Mycobacterium tuberculosis H37Rv Rv3228	Corynebacterium glutamicum ASO19 aroA	Mycobacterium tuberculosis H37Rv Rv3226c	Corynebacterium glutamicum	Mycobacterium tuberculosis H37Rv Rv0336	Mycobacterium tuberculosis sigH
db Match	prf:2214304A		prf:2214304B	pir.F70592	pir:D70592		sp:RR30_SPIOL	gsp:R74093		pir:A70591	pir.F70590	gp:AF114233_1	pir:D70590	GP:AF114233_1	pir:G70506	prf:2515333D
ORF (bp)	678	684	1497	1704	588	156	663	2535	672	504	987	1413	480	123	1110	618
Terminal (nt)	791409	790738	793008	794711	795301	795292	796110	798784	799691	800200	800208	801190	803128	802565	803131	805025
Initial (nt)	790732	791421	791512	793008	794714	795447	795448	796250	799020	799697	801194	802602	802649	802687	804240	804408
SEQ NO.	4340	4341	4342	4343	4344	4345	4346	4347	4348	4349	4350	4351	4352	4353	4354	4355
SEO NO.		841	842	843	844	845	846	847	848	849	850	851	852	853	854	855

EP 1 108 790 A2

_																Т	· · · · · · ·	
7	Function	regulatory protein	hypothetical protein	hypothetical protein	DEAD box ATP-dependent RNA helicase		hypothetical protein	hypothetical protein	ATP-dependent DNA helicase		ATP-dependent DNA helicase		potassium channel	hypothetical protein	DNA helicase II		hypothetical protein	
,	Matched length (a.a.)	84	129	415	458		291	249	1155		1126		302	230	099		280	
	Similarity (%)	96.4	65.1	62.2	64.0		8.69	62.9	48.9		65.7		64.2	58.3	58.8		49.3	
	Identity (%)	78.6	33.3	29.6	37.3		46.4	. 37.0	23.9		41.4		26.2	30.4	32.6		26.8	
Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv3219 whiB1	Mycobacterium tuberculosis H37Rv Rv3217c	Mycobacterium tuberculosis H37Rv Rv3212	Klebsiella pneumoniae CG43 deaD		Mycobacterium tuberculosis H37Rv Rv3207c	Mycobacterium tuberculosis H37Rv Rv3205c	Mycobacterium tuberculosis H37Rv Rv3201c		Mycobacterium tuberculosis H37Rv Rv3201c		Methanococcus jannaschii JAL- 1 MJ0138.1.	Mycobacterium tuberculosis H37Rv Rv3199c	Escherichia coli K12 uvrD		Mycobacterium tuberculosis H37Rv Rv3196	
	db Match	pir.D70596	pir.B70596	pir.E70595	sp:DEAD_KLEPN		pir:H70594	pir:F70594	pir:G70951		pir:G70951		sp:Y13B_METJA	pir.E70951	sp:UVRD_ECOLI		pir:870951	
	ORF (bp)	258	420	1200	1272	225	846	759	3048	780	3219	1332	1005	714	2034	591	816	603
	Terminal (nt)	805535	806737	806740	807946	809510	810394	811163	814217	811386	817422	814210	818523	819236	821287	822669	821290	823391
	Initial (nt)	805792	806318	807939	809217	809286	809549	810405	811170	812165	814204	815541	817519	818523	819254	822079	822105	822789
	SEQ NO. (a.a.)	4356	4357	4358	4359	4360	4361	4362	4363	4364	4365	4366	4367	4368	4369	4370	4371	4372
	SEQ NO.	856	857	858	859	860	861	862	863	864	865	998	867	898	969	870	871	872

EP 1 108 790 A2

5		Function	hypothetical protein	hypothetical protein			hypothetical protein	regulatory protein	ethylene-inducible protein	hypothetical protein	hypothetical protein		alpha-lytic proteinase precursor		DNA-directed DNA polymerase	major secreted protein PS1 protein precursor					monophosphatase
15		Matched length (a.a.)	474	350			1023	463	301	81	201		408		208	363					255
20		Similarity (%)	76.4	74.9			73.5	57.7	89.0	53.0	73.6		44.4		51.4	51.5					74.9
		Identity (%)	42.8	43.4			47.2	34.3	67.4	49.0	40.8		26.7		25.0	27.0					51.8
25	Table 1 (continued)	ns gene	berculosis	berculosis			berculosis	odurans	laticifer er1	K1 APE0247	38 уааЕ		ogenes ATCC		nedia LaBelle- plasmid	glutamicum avum) ATCC					oniger pur3
30	Table 1 (	Homologous gene	Mycobacterium tuberculosis H37Rv Rv3195	Mycobacterium tuberculosis H37Rv Rv3194			Mycobacterium tuberculosis H37Rv Rv3193c	Deinococcus radiodurans DR0840	Hevea brasiliensis laticifer er 1	Aeropyrum pernix K1 APE0247	Bacillus subtilis 168 yaaE		Lysobacter enzymogenes ATCC 29487		Neurospora intermedia LaBelle- 1b mitochondrion plasmid	Corynebacterium glutamicum (Brevibacterium flavum) ATCC 17965 csp1					Streptomyces alboniger pur3
35			ΣI	ΣI			ΣI			×			2 .		ZF						
40		db Match	pir.A70951	pir:H70950			pir.G70950	gp:AE001938_5	sp:ER1_HEVBR	PIR:F72782	sp:YAAE_BACSU		pir.TRYXB4		pir.S03722	sp:CSP1_CORGL		×			prf.2207273H
		ORF (bp)	1446	1050	675	522	2955	1359	951	345	900	363	1062	501	585	1581	429	510	222	309	780
45		Terminal (nt)	822680	825239	825242	825996	829570	829627	831971	831578	832570	832795	834633	835388	835837	838892	839353	840139	840210	840437	841517
50		Initial (nt)	824125	824190	825916	826517	826616	830985	831021	831922	831971	833157	833572	834888	835253	837312	838925	839630	840431	840745	842296
		SEQ NO.	4373	4374	4375	4376	4377	4378	4379	4380	4381	4382	4383	4384	4385	4386	4387	4388	4389	4390	4391
55		SEQ NO.	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	890	891

EP 1 108 790 A2

_						T				$\neg \vdash$	Т			T		$\overline{}$	1	
	Function	myo-inositol monophosphatase	peptide chain release factor 2	cell division ATP-binding protein	hypothetical protein	cell division protein	small protein B (SSRA-binding protein)	hypothetical protein				vibriobactin utilization protein	Fe-regulated protein	hypothetical membrane protein	ferric anguibactin-binding protein precursor	ferrichrome ABC transporter (permease)	ferrichrome ABC transporter (permease)	ferrichrome ABC transporter (ATP-binding protein)
	Matched length (a.a.)	243	359	226	72	301	145	116				272	319	191	325	313	312	250
	Similarity (%)	59.3	88.6	91.2	54.0	74.8	75.9	73.3				52.9	58.3	71.2	61.5	80.8	76.0	82.0
	identity (%)	33.7	68.0	70.4	43.0	40.5	43.5	44.0				26.8	29.5	36.1	27.7	39.3	35.6	48.4
Table 1 (continued)	Homologous gene	Streptomyces flavopersicus spcA	Streptomyces coelicolor A3(2) prfB	Mycobacterium tuberculosis H37Rv Rv3102c ftsE	Aeropyrum pernix K1 APE2061	Mycobacterium tuberculosis H37Rv Rv3101c ftsX	Escherichia coli K12 smpB	Escherichia coli K12 yeaO				Vibrio cholerae OGAWA 395 viuB	Staphylococcus aureus sirA	Mycobacterium leprae MLCB1243.07	Vibrio anguillarum 775 fatB	Bacillus subtilis 168 yclN	Bacillus subtilis 168 yclO	Bacillus subtilis 168 yclP
	db Match	gp:U70376_9	sp:RF2_STRCO	pir.E70919	PIR:G72510	pir.D70919	sp.SMPB_ECOLI	Sp. YEAO_ECOLI				sp:VIUB_VIBCH	prf.2510361A	gp:MLCB1243_5	sp:FATB_VIBAN	pir:B69763	pir.C69763	pir.D69763
	ORF (bp)	819	1104	687	264	006	492	351	537	300	405	825	918	588	1014	666	942	753
	Terminal (nt)	842306	844360	845181	844842	846097	846628	846982	846269	848026	847718	848499	849326	850412	852364	853616	854724	855476
	Initial (nt)	843124	843257	844495	845105	845198	846137	846632	846805	847727	848122	849323	850243	850999	851351	852618	853783	854724
	SEO NO.	4392	4393	4394	4395	4396	4397	4398	4399	4400	4401	4402	4403	4404	4405	4406	4407	4408
	SEQ NO (DNA)	<del></del>	893	894	895	<del></del>	897	898	668	06	901	905	903	904	905	906	907	806

EP 1 108 790 A2

	Function	hypothetical protein	hypothetical protein	kynurenine aminotransferase/glutamine transaminase K		DNA repair helicase	hypothetical protein	hypothetical protein		resuscitation-promoting factor	cold shock protein	hypothetical protein	glutamine cyclotransferase			permease		rRNA(adenosine-2'-0-)- methyltransferase	
	Matched length (a.a.)	48 h	84 h	442 a		613 0	764 h	57 h		198	61 c	159 h	273 g			477 p		319	
	Similarity (%)	72.0	66.0	64.9		62.3	65.2	62.0		64.7	75.4	58.5	67.8			79.3		51.7	
	Identity (%)	0.99	61.0	33.5		30.7	36.1	44.0		39.4	42.6	. 28.3	41.8			43.6		27.9	
Table 1 (continued)	Homologous gene	Chlamydia muridarum Nigg TC0129	Chlamydia pneumoniae	Rattus norvegicus (Rat)		Saccharomyces cerevisiae S288C YIL143C RAD25	Mycobacterium tuberculosis H37Rv Rv0862c	Mycobacterium tuberculosis H37Rv Rv0863		Micrococcus luteus rpf	Lactococcus lactis cspB	Mycobacterium leprae MLCB57.27c	Deinococcus radiodurans DR0112			Streptomyces coelicolor A3(2) SC6C5.09		Streptomyces azureus tsnR	
	db Match	PIR:F81737	GSP: Y35814	pir.S66270		sp:RA25_YEAST	pir.F70815	pir.G70815		prf.2420502A	prf.2320271A	gp:MLCB57_11	gp:AE001874_1			gp:SC6C5_9		sp:TSNR_STRAZ	
	ORF (bp)	147	273	1209	639	1671	2199	219	843	597	381	525	774	669	138	1473	912	828	876
	Terminal (nt)	860078	860473	862752	862753	863396	865119	867571	868630	867803	869318	869379	869918	870721	871660	873210	872016	874040	874069
	Initial (nt)	860224	860745	861544	863391	865066	867317	867353	867788		-	869903	870691	871419	871523	871738	872927	873213	874944
	SEQ NO.	4409	4410	4411	4412	4413	4414	4415	4416	4417	4418	4419	4420	4421	4422	4423	4424	4425	4426
	SEQ.	606	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926

EP 1 108 790 A2

											·									
	Function	hypothetical protein	phosphoserine transaminase	acetyl-coenzyme A carboxylase carboxy transferase subunit beta	hypothetical protein	sodium/proline symporter	•	hypothetical protein	fatty-acid synthase			homoserine O-acetyltransferase			glutaredoxin	dihydrofolate reductase	thymidylate synthase	ammonium transporter	ATP dependent DNA helicase	formamidopyrimidine-DNA glycosidase
	Matched length (a.a.)	316	374	236	103	549		243	3026			335			62	171	261	202	1715	298
	Similarity (%)	55.1	52.9	69.5	80.6	58.1		77.4	83.4			59.7			72.6	62.0	88.9	56.4	68.1	51.0
	Identity (%)	32.6	21.9	36.0	51.5	26.4		49.0	63.1			29.0			43.6	38.0	64.8	32.2	47.4	29.2
Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv0883c	Bacillus circulans ATCC 21783	Escherichia coli K12 accD	Streptomyces coelicolor A3(2) SCIB.08c	Pseudomonas fluorescens		Mycobacterium tuberculosis H37Rv Rv2525c	Corynebacterium ammoniagenes fas			Leptospira meyeri metX			Deinococcus radiodurans DR2085	Mycobacterium avium folA	Escherichia coli K12 thyA	Escherichia coli K12 cysQ	Streptomyces coelicolor A3(2) SC7C7.16c	Synechococcus elongatus naegeli mutM
	db Match	sp:YZ11_MYCTU	pir.S71439	sp:ACCD_ECOL1	gp:SCI8_8	pir.JC2382		pir.A70657	pir:S55505			prf.23173358			gp:AE002044_8	prf:2408256A	sp:TYSY_ECOLI	sp:CYSQ_ECOLI	gp:SC7C7_16	sp:FPG_SYNEN
	ORF (bp)	933	1128	1473	339	1653	816	840	8907	489	186	1047	426	267	237	456	798	756	4560	768
•	Terminal (nt)	874951	875985	879642	881985	883647	884541	884549	894578	895191	895593	895596	896719	897689	897727	897979	898434	899253	904602	905382
	Initial (nt)	875883	877112	881114	881647	881995	883726	885388	885672	894703	895408	896642	897144	897423	897963	898434	899231	900006	900043	904615
	SEQ NO.	4427	4428	4429	4430	4431	4432	4433	4434	4435	4436	4437	4438	4439	4440	4441	4442	4443	4444	4445
	SEQ NO. (DNA)	927	928	† —	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945

EP 1 108 790 A2

~			-																azole- ise	
5					sporter		someas				Se							namide	ninoimid ransfera	
10		Function	hypothetical protein	alkaline phosphatase	integral membrane transporter		glucose-6-phosphate isomease	hypothetical protein		hypothetical protein	ATP-dependent helicase	ABC transporter	ABC transporter		peptidase	hypothetical protein		5'-phosphoribosylglycinamide formyltransferase	5-phosphoribosyl-5-aminoimidazole- 4-carboxamide formyltransferase	citrate lyase (subunit)
15		Matched length (a.a.)	128	196	403	i	557	195		78	763	885	217		236	434		189	525	217
20		Similarity (%)	86.7	71.9	67.0		77.0	.52.3		85.9	73.1	48.6	71.4		73.3	8.09		86.2	87.8	100.0
		Identity (%)	55.5	38.8	33.8		52.4	24.6		59.0	46.1	21.8	43.8		43.6	31.1		64.6	74.5	100.0
25	ontinued)	s gene	erculosis	MG1363 apl	color A3(2)		1101 pgi	erculosis		erculosis	mophilus	icolor A3(2)	8 yvrO		erculosis	erculosis		Z	Į	jlutamicum
30	Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv0870c	Lactococcus lactis MG1363 apl	Streptomyces coelicolor A3(2) SC128.06c		Escherichia coli JM101 pgi	Mycobacterium tuberculosis H37Rv Rv0336		Mycobacterium tuberculosis H37Rv Rv0948c	Bacillus stearothermophilus NCA 1503 pcrA	Streptomyces coelicolor A3(2) SCE25.30	Bacillus subtilis 168 yvrO		Mycobacterium tuberculosis H37Rv Rv0950c	Mycobacterium tuberculosis H37Rv Rv0955		Corynebacterium ammoniagenes purN	Corynebacterium ammoniagenes purH	Corynebacterium glutamicum ATCC 13032 citE
35			£Ϋ́	عا	% स		Es	ΣÏ				ळ ळ	B		ŽΪ		_	ਤੇ ਨ	उ ह	
40		db Match	pir.F70816	SP.APL_LACLA	pir.T36776		pir.NUEC	pir.G70506		sp:YT26_MYCTU	sp:PCRA_BACST	gp:SCE25_30	prf.2420410P		pir:D70716	sp:YT19_MYCTU		gp:AB003159_2	gp:AB003159_3	gp:CGL133719_3
		ORF (bp)	408	009	1173	717	1620	1176	381	309	2289	2223	999	507	711	1425	228	627	1560	819
45		Terminal (nt)	905796	905792	906559	909328	907759	909521	911223	910855	913514	913477	915699	916368	916970	919352	917827	919956	921526	922412
50		Initial (nt)	905389	906391	907731	908612	909378	910696	910843	911163	911226	915699	916364	916874	917680	917928	918054	919330	919967	921594
		SEQ NO.	4446	4447	4448	4449	4450	4451	4452	4453	4454	4455	4456	4457	4458	4459	4460	4461	4462	4463
<i>55</i>		SEQ NO (DNA)		947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	362	963

ħ

EP 1 108 790 A2

5		Function	repressor of the high-affinity (methyl) ammonium uptake system	hypothetical protein		30S ribosomal protein S18	30S ribosomal protein S14	50S ribosomal protein L33	50S ribosomal protein L28	transporter (sulfate transporter)	Zn/Co transport repressor	50S ribosomal protein L31	50S ribosomal protein L32		copper-inducible two-component regulator	two-component system sensor	proteinase DO precursor	molybdopterin biosynthesis cnx1 protein (molybdenum cofactor biosynthesis enzyme cnx1)		large-conductance mechanosensitive channel	hypothetical protein	5-formyltetrahydrofolate cyclo-ligase
15		Matched length (a.a.)	222	109		29	100	49	77	529	. 80	78	55		227	484	406	188		131	210	191
20		Similarity (%)	100.0	100.0		76.1	80.0	83.7	81.8	71.1	77.5	65.4	78.2		73.6	60.1	59.9	54.3		77.1	60.0	59.7
		Identity (%)	100.0	100.0		52.2	54.0	55.1	52.0	34.4	37.5	37.2	60.0		48.0	24.4	33.3	27.7		50.4	28.6	25.1
25	Table 1 (continued)	ns gene	glutamicum R	glutamicum		doxa rps18	12 rpsN	12 rpmG	12 rpmB	38 yvdB	ureus zntR	reyi rpmE	licolor A3(2)		ringae copR	12 baeS	.12 htrA	ına CV cnx1		berculosis nscl.	berculosis	THFS
30	Table 1 (	Homologous gene	Corynebacterium glutamicum ATCC 13032 amtR	Corynebacterium glutamicum ATCC 13032 yjcC		Cyanophora paradoxa rps18	Escherichia coli K12 rpsN	Escherichia coli K12 rpmG	Escherichia coli K12 rpmB	Bacillus subtilis 168 yvdB	Staphylococcus aureus zntR	Haemophilus ducreyi rpmE	Streptomyces coelicolor A3(2) SCF51A.14		Pseudomonas syringae copR	Escherichia coli K12 baeS	Escherichia coli K12 htrA	Arabidopsis thaliana CV cnx1		Mycobacterium tuberculosis H37Rv Rv0985c msct.	Mycobacterium tuberculosis H37Rv Rv0990	Homo sapiens MTHFS
35		ے	2.						ш	8					_		w ·				<u> </u>	
40		db Match	gp:CGL133719_	gp:CGL133719_1		sp:RR18_CYAPA	Sp.RS14_ECOLI	sp:RL33_ECOL	pir.R5EC28	pir.B70033	prf:2420312A	Sp:RL31_HAEDU	gp:SC51A_14		sp:COPR_PSESM	sp:BAES_ECOLI	pir.S45229	sp:CNX1_ARATH		sp:MSCL_MYCTU	pir.A70601	pir.JC4389
		ORF (bp)	999	327	321	249	303	162	234	1611	312	264	171	447	969	1365	1239	585	198	405	651	570
45		Terminal (nt)	922396	923138	923981	924159	924425	924734	924901	925325	926931	927737	927922	927339	928812	930248	931648	932290	932487	932570	933060	933733
50		Initial (nt)	923061	923464	923661	924407	924727	924895	925134	926935	927242	927474	927752	927785	928117	928884	930410	931706	932290	932974	933710	934302
		SEQ NO. (a.a.)	4464	4465	4466	4467	4468	4469	4470	4471	4472	4473	4474	4475	4476	4477	4478	4479	4480	4481	4482	4483
55		SEQ NO.	964	965	996	296	996	696	970	971	972	973	974	975	976	977	978	979	980	981	982	983

## EP 1 108 790 A2

																			$\overline{}$
10		Function	UTP-glucose-1-phosphate uridylyltransferase	molybdopterin biosynthesis protein	ribosomal-protein-alanine N- acetyltransferase	hypothetical membrane protein	cyanate transport protein		hypothetical membrane protein	hypothetical membrane protein	cyclomaltodextrinase	hypothetical membrane protein	hypothetical protein	methionyl-tRNA synthetase	ATP-dependent DNA helicase	hypothetical protein	hypothetical protein		transposase
15		Matched length (a.a.)	296	390	193	367	380		137	225	444	488	272	615	741	210	363		94
20		Similarity (%)	68.9	62.6	54.9	54.8	62.4		9.09	59.6	53.6	75.2	78.3	66.7	49.0	53.3	59.0		59.6
		Identity (%)	42.2	31.8	29.0	30.3	26.6		32.1	25.3	26.8	43.0	54.0	33.8	26.2	27.6	30.0		33.0
25	Table 1 (continued)	us gene	npestris	inovorans	12 rimJ	berculosis	12 cynX		lenzae Rd	Ibercutosis	ıs E-244	berculosis	iberculosis	n cum Delta H	ecQ	ກ cum Delta H	68 yxaG		cium
30	Table 1 (	Homologous gene	Xanthomonas campestris	Arthrobacter nicotinovorans moeA	Escherichia coli K12 rimJ	Mycobacterium tuberculosis H37Rv Rv0996	Escherichia coli K12 cynX		Haemophilus influenzae Rd H11602	Mycobacterium tuberculosis H37Rv Rv0093c	Bacillus sphaericus E-244 CDase	Mycobacterium tuberculosis H37Rv	Mycobacterium tuberculosis H37Rv Rv1003	Methanobacterium thermoautotrophicum Delta H MTH587 metG	Escherichia coli recQ	Methanobacterium thermoautotrophicum Delta H MTH796	Bacillus subtilis 168 yxaG		Enterococcus faecium
35			×	ΚE	Ш	ΣI			II		l	21			au	2=2			
40		db Match	pir.JC4985	prf.2403296B	sp:RIMJ_ECOLI	pir:G70601	Sp.CYNX_ECOLI		sp:YG02_HAEIN	sp:Y05C_MYCTU	sp:CDAS_BACSH	pir.E70602	sp:Y19J_MYCTU	SP.SYM_METTH	prf: 1306383A	pir: <b>B69</b> 206	sp:YXAG_BACSU		gp:AF029727_1
		ORF (bp)	897	1257	099	1020	1200	1419	405	714	1167	1560	825	1830	2049	633	1158	531	294
45		Terminal (nt)	935319	936607	937274	938401	939626	937799	940090	940754	941925	942381	944833	948669	950839	950828	951834	953043	954266
50		Initial (nt)	934423	935351	936615	937382	938427	939217	939686	940041	940759	943940	944009	946840	948791	951460	952991	953573	953973
		SEQ NO.	4484	4485	4486	4487	4488	4489	4490	4491	4492	4493	4494	4495	4496	4497	4498	4499	4500
55		SEQ NO.		985	986	286	988	686	066	991	992	993	994	982	966	1997	968	666	1000

5					ase	hyltransferase				or	orotein				ansferase	sphate kinase			-		
10	Function	transposase	transposase subunit		D-lactate dehydrogenase	site-specific DNA-methyltransferase		transposase	transposase	transcriptional regulator	cadmium resistance protein		hypothetical protein	hypothetical protein	dimethyladenosine transferase	isopentenyl monophosphate kinase		ABC transporter	pyridoxine kinase	hypothetical protein	hypothetical protein
15	Matched length (a.a.)	139	112		565	231		94	139	91	205		263	362	265	315		478	242	159	108
20	Similarity (%)	9'.29	88.4		75.6	62.8		59.6	67.6	84.6	66.8		70.7	63.5	65.3	67.0		82.8	67.4	· 58.5	78.7
	identify (%)	41.7	73.2		46.4	30.8		33.0	41.7	62.6	31.7		46.4	34.8	34.3	42.5		65.5	40.1	27.0	45.4
25 (panui	ene		tnpA			OK8				ulosis	s cadD		ulosis	ulosis	sgA	ulosis		ythraea	dxK	ulosis	or A3(2)
& Table 1 (continued)	Homologous gene	Escherichia coli K12	Brevibacterium linens tnpA		Escherichia coli dld	Klebsiella pneumoniae OK8 kpnIM		Enterococcus faecium	Escherichia coli K12	Mycobacterium tuberculosis H37Rv Rv1994c	Staphylococcus aureus cadD		Mycobacterium tuberculosis H37Rv Rv1008	Mycobacterium tuberculosis H37Rv Rv1009 rpf	Escherichia coli K12 ksgA	Mycobacterium tuberculosis H37Rv Rv1011		Saccharopolyspora erythraea ertX	Escherichia coli K12 pdxK	Mycobacterium tuberculosis H37Rv Rv2874	Streptomyces coelicolor A3(2) SCF1.02
35		ш						-,					S T	~ 1	ECOLI	<u> </u>		மை		•	000
40	db Match	pir.TQEC13	gp:AF052055_1		prf:2014253AE	sp:MTK1_KLEPN	:	gp:AF029727	pir.TQEC13	sp:YJ94_MYCTU	prf:2514367A		pir.C70603	pir.D70603	sp:KSGA_EC	pir.F70603		pir.S47441	sp:PDXK_ECOLI	sp:YX05_MYCTU	gp:SCF1_2
	ORF (bp)	477	414	864	1713	840	219	294	477	357	621	342	831	1071	879	933	642	1833	792	480	321
45	Terminal (nt)	954753	955354	956774	955686	957844	959185	960374	960861	961653	962249	961321	963639	964934	965852	966784	965950	968660	969458	969461	970349
50	Initial (nt)	954277	954941	955911	957398	958683	959403	960081	960385	961297	961629	961662	962809	963864	964974	965852	966591	966828	968667	969940	970029
	SEQ NO. (a.a.)	4501	4502	4503	4504	4505	4506	4507	4508	4509	4510	4511	4512	4513	4514	4515	4516	4517	4518	4519	4520
55	SEQ NO.	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010	1011	1012	1013	1014	1015	1016	1017	1018	1019	1020

EP 1 108 790 A2

	- (						Т	ŤТ			· · · · · ·		1	$\neg$				$\neg$	
5		u,		-						in PS1 protein	itor (tetR	protein	ie.2- ne			:		e factor 3	protein
10		Function	hypothetical protein	regulator	hypothetical protein	enoyl-CoA hydratase				major secreted protein PS1 protein precursor	transcriptional regulator (tetR family )	membrane transport protein	S-adenosylmethionine:2- demethylmenaquinone methyltransferase		hypothetical protein	hypothetical protein		peptide-chain-release factor 3	amide-urea transport protein
15		Matched length (a.a.)	107	261	276	337				440	100	802	157		121	482		546	404
20		Similarity (%)	69.2	88.1	59.1	6.07				56.8	0.07	70.0	75.8		63.6	48.3		68.0	72.8
		Identity (%)	35.5	64.8	27.2	35.6		-		27.7	44.0	42.6	38.2		29.8	24.9		39.2	42.8
25	Table 1 (continued)	us gene	licolor A3(2)	licolor A3(2)	8 ухеН	berculosis				glutamicum avum) ATCC	licolor A3(2)	licolor A3(2)	enzae Rd		tidis NMA1953	berculosis		12 prfC	hylotrophus
30	Table 1 (c	Homologous gene	Streptomyces coelicolor A3(2) SCF1.02	Streptomyces coelicolor A3(2) SCJ1.15	Bacillus subtilis 168 yxeH	Mycobacterium tuberculosis H37Rv echA9				Corynebacterium glutamicum (Brevibacterium flavum) ATCC 17965 csp1	Streptomyces coelicolor A3(2) SCF56.06	Streptomyces coelicolor A3(2) SCE87.17c	Haemophilus influenzae Rd H10508 menG		Neisseria meningitidis NMA1953	Mycobacterium tuberculosis H37Rv Rv1128c		Escherichia coli K12 prfC	Methylophilus methylotrophus fmdD
40		db Match	gp:SCF1_2	gp:SCJ1_15	sp:YXEH_BACSU E	pir:E70893				sp.CSP1_CORGL (	gp:SCF56_6	gp:SCE87_17	sp.MENG_HAEIN		gp:NMA6Z2491_21	pir.A70539		pir.159305	prf.2406311A
		ORF (bp)	321	096	792	1017	654	111	1212	1386	579	2373	498	999	381	1551	936	1647	1269
45 .		Terminal (nt)	970738	971823	972244	974155	973304	974962	974965	977734	977800	978368	981490	982287	982294	984650	985845	984864	988007
50		Initial (nt)	970418	970864	973035	973139	973957	974186	976176	976349	978378	980740	980993	981622	982674	983100	984910	986510	986739
		SEQ NO. (a.a.)	4521	4522	4523	4524	4525	4526	4527	4528	4529	4530	4531	4532	4533	4534	4535	4536	4537
55		SEQ NO. (DNA)	1021	1022	1023	1024	1025	1026	1027	1028	1029	1030	1031	1032	1033	1034	1035	1036	1037

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	•

,						$\neg \tau$	$\overline{}$				1							
	Function	amide urea transport protein	amide-urea transport protein	high-affinity branched-chain amino acid transport ATP-binding protein	high-affinity branched-chain amino acid transport ATP-binding protein	peptidyl-tRNA hydrolase	2-nitropropane dioxygenase	glyceraldehyde-3-phosphate dehydrogenase	polypeptides predicted to be useful antigens for vaccines and diagnostics	peptidyl-tRNA hydrolase	50S ribosomal protein L25	lactoylgiutathione lyase	DNA alkylation repair enzyme	ribose-phosphate pyrophosphokinase	UDP-N-acetylglucosamine pyrophosphorylase		sufl protein precursor	nodulation ATP-binding protein I
	Matched length (a.a.)	7.7	234	253	236	187	361	342	51	174	194	143	208	316	452		909	310
	Similarity (%)	61.0	68.0	0.07	69.1	9.07	54.0	72.8	61.0	63.2	65.0	54.6	62.5	79.1	71.9	-	61.7	64.8
	Identity (%)	40.8	34.6	37.9	35.2	39.0	25.2	39.5	54.0	38.5	47.0	28.7	38.9	44.0	42.0		30.8	35.8
Table 1 (continued)	Homologous gene	Methylophilus methylotrophus fmdE	Methylophilus methylotrophus fmdF	Pseudomonas aeruginosa PAO braF	Pseudomonas aeruginosa PAO braG	Escherichia coli K12 pth	Williopsis mrakii IFO 0895	Streptomyćes roseofulvus gap	Neisseria meningitidis	Escherichia coli K12 pth	Mycobacterium tuberculosis H37Rv rplY	Salmonella typhimurium D21 gloA	Bacillus cereus ATCC 10987 aIkD	Bacillus subtilis prs	Bacillus subtilis gcaD		Escherichia coli K12 sufl	Rhizobium sp. N33 nodl
	db Match	prf:2406311B	prf.2406311C	sp.BRAF_PSEAE	sp:BRAG_PSEAE	Sp:PTH_ECOLI	sp:2NPD_WILMR	sp.G3P_ZYMMO	GSP:Y75094	sp:PTH_ECOLI	pir:B70622	sp:LGUL_SALTY	prf:2516401BW	sp:KPRS_BACCL	pir:S66080		sp:SUFI_ECOLI	sp:NODI_RHIS3
	ORF (bp)	882	1077	726	669	612	1023	1065	369	531	900	429	624	975	1455	1227	1533	918
	Terminal (nt)	988904	989980	990705	991414	991417	993080	994613	994106	994845	995527	996830	996833	997466	998455	1000016	1002864	1003930
	Initiat (nt)	988023	988904	989980	990716	992028	992058	993549	994474	995375	996126	996402	997456	998440	606666	1001242	1001332	1003013
	SEQ NO.	4538	4539	4540	4541	4542	4543	4544	4545	4546	4547	4548	4549	4550	4551	4552	4553	4554
	A O E	038	039	040	041	042	043	044	045	046	047	048	049	020	051	052	053	054

																							_
5		ç	ne protein	m sensor	criptional /)		ne protein				speptidase					fragment	3 TnpB)				ator (TetR-	coupling protein	
10		Function	hypothetical membrane protein	two-component system sensor histidine kinase	two component transcriptional regulator (luxR family)		hypothetical membrane protein	ABC transporter		ABC transporter	gamma-glutamyltranspeptidase precursor					transposase protein fragment	transposase (IS1628 TnpB)				transcriptional regulator (TetR- family)	transcription/repair-coupling protein	
15		Matched length (a.a.)	272	459	202		349	535		573	999					37	236				183	1217	
20		Similarity (%)	63.2	48.4	67.3		64.5	57.0		74.0	58.6					72.0	100.0				59.6	65.1	
		Identity (%)	30.2	24.6	36.6		31.5	28.6		44.0	32.4					64.0	9.66				23.0	36.2	
25	(pant	ne	DRF2	Bq	s dnrN		r A3(2)	ens strV		atis exiT	*					micum	micum 1 tnpB				!		
30	Table 1 (continued)	Homologous gene	Streptomyces lividans ORF2	Escherichia coli K12 uhpB	Streptomyces peucetius dnrN		Streptomyces coelicolor A3(2) SCF15.07	Streptomyces glaucescens strV		Mycobacterium smegmatis exiT	Escherichia coli K12 ggt					Corynebacterium glutamicum TnpNC	Corynebacterium glutamicum 22243 R-plasmid pAG1 tnpB				Escherichia coli tetR	Escherichia coli mfd	
35			क		S S		<i>I I I I I I I I I I</i>	S		Σ													
40		db Match	pir.JN0850	sp:UHPB_ECOLI	prf.2107255A		gp:SCF15_7	pir.S65587		pir.T14180	sp:GGT_ECOLI					GPU.AF164956_23	gp:AF121000_8				sp:TETC_ECOLI	Sp:MFD_ECOLI	
		ORF (bp)	831	1	609	204	1155	1440	153	1734	1965	249	519	192	606	243	708	462	597	312	651	3627	1224
45		Terminal (nt)	1004783	1006085	1006697	1006734	1008152	1010061	1008534	1011790	1011797	1014264	1014343	1015116	1016560	1015450	1015145	1017018	1017274	1018393	1019066	1022716	1019390
50		Initial (nt)	1003953	1004829	1006089	1006937	1006998	1008622	1008686	1010057	1013761	1014016	1014861	1014925	1015652	1015692	1015852	1016557	1017870	1018082	1018416	1019090	1020613
		SEQ NO.	4555	4556	4557	4558	4559	4560	4561	4562	4563	4564	4565	4566	4567	4568	4569	4570	4571	4572	4573	4574	4575
		O O	55	356	750	28	)59	186	190	362	063	98	90	990	190	890	690	070	071	072	073	074	075

		F			ī		<del>-</del> T	Т	Т	Т			7					Т	$\neg$
5		uo uo	des predicted to or vaccines and	Like ATP-		ane protein						iglycerate spho-D- se)				osphatase or e		156	
10		Function	Neisserial polypeptides predicted to be useful antigens for vaccines and diagnostics	multidrug resistance-like ATP- binding protein, ABC-type transport protein	ABC transporter	hypothetical membrane protein		hypothetical protein			IpqU protein	enolase (2-phosphoglycerate dehydratase)(2-phospho-D- glycerate hydro-lyase)	hypothetical protein	hypothetical protein	hypothetical protein	guanosine pentaphosphatase or exopolyphosphatase		threonine dehydratase	
15		Matched length (a.a.)	76	632	574	368		183	-		241	422	. 41	191	153	329		314	
20		Similarity (%)	0.69	62.7	81.9	100.0		57.4			68.9	86.0	58.0	55.0	77.8	55.0		64.7	
		Identity (%)	48.0	31.3	50.2	100.0		33.4			46.5	64.5	68.0	31.9	59.5	25.2		30.3	
25	ntinued)	gene	396	8	erculosis	utamicum		Z			erculosis U		(1 APE2459	erculosis	erculosis	pA .		<b>e</b>	
30	Table 1 (continued)	Homologous gene	Neisseria gonorrhoeae	Escherichia coli mdlB	Mycobacterium tuberculosis H37Rv Rv1273c	Corynebacterium glutamicum ATCC 13032 orf3		Bacillus subtilis yabN			Mycobacterium tuberculosis H37Rv Rv1022 IpqU	Bacillus subtilis eno	Aeropyrum pernix K1 APE2459	Mycobacterium tuberculosis H37Rv Rv1024	Mycobacterium tuberculosis H37Rv Rv1025	Escherichia coli gppA		Escherichia coli tdcB	
35		db Match	SP.Y75301	sp:MDLB_ECOLI	sp:YC73_MYCTU	YLI3_CORGL		sp:YABN_BACSU			ir.A70623	p:ENO_BACSU	IR:872477	ir.C70623	ir.D70623	p:GPPA_ECOLI		p.THD2_ECOLI	
40			9		<del></del>	S			-		ā.	S	<u>n</u>	-	α.	S	-	S	10
		ORF (bp)	228	1968	1731	2382	297	585	426	378	786	1275	144	540	546	963	984	930	195
45		Terminal (nt)	1021078	1022699	1024666	1026505	1032181	1032780	1032760	1033269	1034739	1036223	1036016	1036855	1037445	1038410	1036498	1038721	1039977
50		Initial (nt)	1021305	1024666	1026396	1028886	1031885	1032196	1033185	1033646	1033954	1034949	1036159	1036316	1036900	1037448	1037481	1039650	1039783
		SEQ.	4576	4577	4578	4579	4580	4581	4582	4583	4584	4585	4586	4587	4588	4589	4590	4591	4592
55		SEQ		1077	1078	1079	1080	1081	1082	1083	1084	1085	1086	1087	1088	1089	1090	1091	1092

5
10
15
20
25
30
35
40
<b>45</b>
50

Table 1 (continued)

		Γ							Matchad	
SEQ Initial Terminal ORF NO. (nt) (nt) (bp)	Terminal (nt)		ORF (bp)		db Match	Homologous gene	Identity (%)	Similarity (%)	length (a.a.)	Function
4593 1039996 1040325 330	1040325	<del> </del> —	330							
4594 1040494 1040682 189	1040682		189		pir:B72287	Thermotoga maritima MSB8	46.3	74.1	56	hypothetical protein
4595 1040925 1041917 993	1041917	<del> </del>	993		sp.RHAR_ECOLI	Escherichia coli rhaR	24.8	55.8	242	transcription activator of L-rhamnose operon
4596 1042027 1042842 816	1042842	<del> </del> -	816		pir:F70893	Mycobacterium tuberculosis H37Rv Rv1072	57.8	80.1	282	hypothetical protein
4597 1043236 1042850 387	1042850		387							
4598 1043747 1043298 450	1043747 1043298	-	450		gp:SCF55_39	Streptomyces coelicolor A3(2) SCF55.39	30.0	57.1	140	hypothetical protein
4599 1044295 1043774 522	1044295 1043774 522	522	<del> </del>		Sp.GREA_ECOLI	Escherichia coli greA	35.0	60.1	143	transcription elongation factor
4600 1044959 1044477 483 p	1044959 1044477 483	483		<u> </u>	pir.G70894	Mycobacterium tuberculosis H37Rv Rv1081c	34.3	72.1	140	hypothetical protein
4601 1045158 1046030 873 p	1046030 873	873	$\vdash$	۵	pir.S44952	Streptomyces lincolnensis ImbE	31.7	56.3	300	lincomycin-production
4602 1046073 1046390 318	1046390		318							
4603 1046610 1047707 1098	1046610 1047707 1098	1098			sp:AROG_CORGL	Corynebacterium glutamicum aroG	99.2	99.5	367	3-deoxy-D-arabino-heptulosonate-7-phosphate synthase
4604 1047452 1046820 633	1047452 1046820		633							
4605 1047827 1048501 675	1047827 1048501 675	675			sp:YARF_CORGL	Corynebacterium glutamicum CCRC18310	0.96	97.3	26	hypothetical protein or undecaprenyl pyrophosphate synthetase
4606 1048356 1048529 174	1048356 1048529		174		SP:YARF_CORGL	Corynebacterium glutamicum (Brevibacterium flavum)	100.0	100.0	28	hypothetical protein
4607 1048525 1049043 519	1049043		519							
4608 1049385 1049068 318	1049385 1049068	_	318							
4609 1050362 1049427 936	1050362 1049427		936		sp:coaa_Ecoul	Escherichia coli coaA	53.9	79.9	308	pantothenate kinase
4610 1050624 1051925 1302	1050624 1051925		1302		gsp:R97745	Brevibacterium flavum MJ-233 glyA	99.5	100.0	434	serine hydroxymethyl transferase
4611 1052021 1053880 1860	1052021 1053880		1860		sp:PABS_STRGR	Streptomyces griseus pabS	47.6	70.1	969	p-aminobenzoic acid synthase
4612 1053880 1054602 - 723	1053880 1054602	1054602 -								

		_											_				· 		т					<del></del> -	$\neg$
5			Function			phosphinothricin resistance protin	hypothetical protein		hypothetical protein	lactam utilization protein	hypothetical membrane protein			transcriptional regulator		fumarate hydratase precursor	NADH-dependent FMN oxydoreductase			reductase	dibenzothiophene desulfurization enzyme A	dibenzothiophene desulfurization enzyme C (DBT sulfur dioxygenase)	dibenzothiophene desulfurization enzyme C (DBT sulfur dioxygenase)		
15			Matched length (a.a.)			165 pt	300 h	$\neg$	225 hy	276 la	165 hy			204 tr		456 fu	159 N			184 re	443 di	372 d	391 d		
20			Similarity (%)			58.8	29.0		57.8	52.2	81.2			63.2		79.4	65.4			81.0	67.7	51.3	61.6		
			Identity (%)			30.3	30.3		37.8	30.8	40.6			26.0		52.0	32.7			55.4	39.1	25.8	28.9		
25	. :	nued)	ene			c <sub>R</sub>				mB						t) fumH	olis			or A3(2)	S8 soxA	IGTS8 soxC	S8 soxC		
30		Table 1 (continued)	Homologous gene			Alcaligenes faecalis ptcR	Escherichia coli ybgK		Escherichia coli ybgJ	Emericella nidulans lamB	Bacillus subtilis ycsH			Bacillus subtilis ydhC		Rattus norvegicus (Rat) fumH	Rhodococcus erythropolis IGTS8 dszD			Streptomyces coelicolor A3(2) StAH10.16	Rhodococcus sp. IGTS8 soxA	Rhodococcus sp. IGT	Rhodococcus sp. IGTS8 soxC		
35						ΑK	Es		Es	ä	Ва			a		Ra	윤일			क्र क्र		<del>                                     </del>			
40			db Match			gp:A01504_1	sp:YBGK_ECOLI		sp.YBGJ_ECOLI	SP:LAMB_EMEN!	SP:YCSH_BACSU			sp.YDHC_BACSU		Sp:FUMH_RAT	gp:AF048979_1			gp:SCAH10_16	sp:SOXA_RHOSO	sp:SOXC_RHOSO	sp:SOXC_RHOSO		
			ORF (bp)	864	393	537	879	1056	699	756	591	672	603	681	1278	1419	489	261	447	564	1488	1080	1197	780	690
45			Terminal (nt)	1055722	1054640	1056319	1056322	1058628	1057200	1057843	1058624	1059889	1059962	1060792	1062146	1062211	1064424	1064478	1064754	1065304	1067570	1068649	1069845	1068913	1069119
50			Initial (nt)	1054859	1055032	1055783	1057200	1057573	1057868	1058598	1059214	1059218	1059360	1060112	1060869	1063629	1063936	1064738	1065200	1065867	1066083	1067570	1068649	1069692	1069808
			SEQ NO.	4613	4614	4615	4616	4617	4618	4619	4620	4621	4622	4623	4624	4625	4626	4627	4628	4629	4630	4631	4632	4633	4634
55			SEQ NO.	1113	1114	1115	1116	1117	1118	1119	1120	1121	1122	1123	1124	1125	1126	1127	1128	1129	1130	1131	1132	1133	1134

5
10
15
20
25
30
35
40
45
50

	Function	FMNH2-dependent aliphatic sulfonate monooxygenase	glycerol metabolism	hypothetical protein	hypothetical protein		transmembrane efflux protein	exodeoxyribonuclease small subunit	exodeoxyribonuclease large subunit	penicillin tolerance	polypeptides predicted to be useful antigens for vaccines and diagnostics		permease		sodium-dependent proline transporter	major secreted protein PS1 protein precursor	GTP-binding protein	virulence-associated protein	ornithine carbamoyltransferase	hypothetical protein
	Matched length (a.a.)	397	325	211	227		82	62	466	311	131		338		252	412	361	75	301	143
	Similarity (%)	73.1	75.7	56.4	66.1		78.1	67.7	55.6	78.8	47.0		63.9		61.4	60.0	88.6	0.08	58.8	6.69
	Identity (%)	45.3	44.3	27.5	31.3		36.6	40.3	30.0	50.2	33.0		26.3		30.3	29.9	70.1	6.73	29.6	39.2
Table 1 (continued)	Homologous gene	Escherichia coli K12 ssuD	Escherichia coli K12 glpX	Mycobacterium tuberculosis H37Rv Rv1100	Bacillus subtilis ywmD		Streptomyces coelicolor A3(2) SCH24.37	Escherichia coli K12 MG1655 xseB	Escherichia coli K12 MG1655 xseA	Escherichia coli K12 lytB	Neisseria gonorrhoeae		Escherichia coli K12 perM		Rattus norvegicus (Rat) SLC6A7 ntpR	Corynebacterium glutamicum (Brevibacterium flavum) ATCC 17965 csp1	Bacillus subtilis yyaF	Dichelobacter nodosus intA	Pseudomonas aeruginosa argF	Bacillus subtilis 168 ykkB
	db Match	gp:ECO237695_3	sp:GLPX_ECOLI	pir:B70897	pir:H70062		gp:SCH24_37	sp:EX7S_ECOLI	sp:EX7L_ECOLI	Sp:LYTB_ECOLI	GSP:Y75421		Sp:PERM_ECOLI		sp:NTPR_RAT	sp:CSP1_CORGL	sp:YYAF_BACSU	sp:VAPI_BACNO	sp:OTCA_PSEAE	sp:YKKB_BACSU
	ORF (bp)	1176	963	570	1902	285	225	243	1251	975	429	828	1320	180	1737	1233	1083	297	822	501
	Terminal (nt)	1071134	1071479	1073245	1073340	1075641	1075329	1075667	1075933	1078271	1077306	1078319	1079221	1080786	1080972	1082951	1085462	1086087	1086917	1087044
	Initial (nt)	1069959	1072441	1072676	1075241	1075357	1075553	1075909	1077183	1077297	1077734	1079146	1080540	1080965	1082708	1084183	1084380	1085791	1086096	1087544
	SEQ NO. (a.a.)	4635	4636	4637	4638	4639	4640	4641	4642	4643	4644	4645	4646	4647	4648	4649	4650	4651	4652	4653
:	SEQ NO. (DNA)	1135	1136	1137	1138	1139	1140	1141	1142	1143	1144	1145	1146	1147	1148	1149	1150	1151	1152	1153

						$\neg$	Т	$\Box$											ပ
5		Function	9-cis retinol dehydrogenase or oxidoreductase	transposase/integrase (IS110)	hypothetical membrane protein	N-acetylglucosaminyltransferase			transposase (insertion sequence IS31831)	transposase	transposase				oxidoreductase or morpyine-6- dehydrogenase (naloxone reductase)	4-carboxymuconolactone decarboxlyase			frenolicin gene cluster protein involved in frenolicin biosynthetic
15		Matched length (a.a.)	198	396	1153	259			6	125	48				264	108			146
20		Similarity (%)	9.09	73.0	52.2	47.1			93.8	94.4	95.8				66.3	63.9			66.4
		Identity (%)	33.8	42.2	23.0	22.8			82.5	79.2	87.5				37.5	33.3			34.9
25	Table 1 (continued)	ans gene	H4	licolor	12 yegE	i nodC			glutamicum	glutamicum actofermentum)	glutamicum actofermentum)				tida M10 norA	coaceticus			eofulvus frnS
30	Table 1 (	Homologous gene	Mus musculus RDH4	Streptomyces coelicolor SC3C8.10	Escherichia coli K12 yegE	Rhizobium meliloti nodC			Corynebacterium glutamicum ATCC 31831	Corynebacterium glutamicum (Brevibacterium lactofermentum) ATCC 13869	Corynebacterium glutamicum (Brevibacterium lactofermentum) ATCC 13869				Pseudomonas putida M10 norA	Acinetobacter calcoaceticus dc4c			Streptomyces roseofulvus frnS
35				-					UQ	USA	0 0 4								
40		db Match	gp:AF013288_1	sp:YIS1_STRCO	sp:YEGE_ECOLI	SP:NODC_RHIME			pir.S43613	pir.JC4742	pir.JC4742				sp:MORA_PSEPU	sp:DC4C_ACICA			gp:AF058302_19
		ORF (bp)	630	1206	3042	765	219	333	291	375	144	141	366	498	843	321	663	195	654
45		Terminal (nt)	1087664	1088535	1093216	1094693	1094911	1095384	1095387	1095719	1096188	1096331	1096746	1097726	1098592	1098929	1099750	1099015	1099115
50		Initial (nt)	1088293	1089740	1090175	1093929	1094693	1095052	1095677	1096093	1096331	1096471	1097111	1097229	1097750	1098609	1099088	1099209	1099768
		SEQ NO.	4654	4655	4656	4657	4658	4659	4660	4661	4662	4663	4664	4665	4666	4667	4668	4669	
55		SEQ NO.		1155	1156	1157	1158	1159	1160	1161	1162	1163	1164	1165	1166	1167	1168	1169	1170

r	—т		Т	$\neg$	$\neg \tau$		$\neg \tau$									$\neg$		
	Function	biotin carboxylase				-		hypothetical protein	magnesium chelatase subunit	2,3-PDG dependent phosphoglycerate mutase	hypothetical protein	carboxyphosphonoenolpyruvate phosphonomutase	tyrosin resistance ATP-binding protein	hypothetical protein	alkylphosphonate uptake protein	transcriptional regulator	multi-drug resistance efflux pump	transposase (insertion sequence IS31831)
	Matched length (a.a.)	563						655	329	160	262	248	593	136	111	134	367	436
	Similarity (%)	78.5						80.3	52.6	62.5	60.7	59.3	54.1	6.99	82.0	62.7	59.4	99.8
	Identity (%)	48.1						57.9	27.7	33.8	38.2	29.4	31.7	29.4	55.0	32.1	22.6	99.5
Table 1 (continued)	Homologous gene	Synechacoccus sp. PCC 7942 accC						Mycobacterium tuberculosis H37Rv Rv0959	Rhodobacter sphaeroides ATCC 17023 bchl	Amycolatopsis methanolica pgm	Mycobacterium tuberculosis H37Rv Rv2133c	Streptomyces hygroscopicus SF1293 BcpA	Streptomyces fradiae ttrC	Mycobacterium tuberculosis H37Rv Rv2923c	Escherichia coli K12 MG1655 phnA	Bacil∤us subtilis 168 yxa□	Streptococcus pneumoniae pmrA	Corynebacterium glutamicum (Brevibacterium lactofermentum) ATCC 31831
	db Match	gp:SPU59234_3						sp:YT15_MYCTU	sp:BCHI_RHOSH	gp:AMU73808_1	pir.A70577	gp:STMBCPA_1	sp:TLRC_STRFR	sp:Y06C_MYCTU	sp:PHNA_ECOL!	sp:YXAD_BACSU	gp:SPN7367_1	pir.S43613
	ORF (bp)	1737	597	498	345	153	639	1956	1296	642	705	762	1641	396	342	474	1218	1308
	Terminal (nt)	1101653	1102639	1103192	1103524	1104103	1105561	1104103	1106086	1108201	1108905	1109754	1111432	1111425	1112230	1112484	1114319	1115793
	Initial (nt)	1099917	1102043	1102695	1103180	1103951	1104923	1106058	1107381	1107560	1108201	1108993	1109792	1111820	1111889	1112957	1113102	1114486
	SEQ NO.	<del>!</del>	4672	4673	4674	4675	4676	4677	4678	4679	4680	4681	4682	4683	4684	4685	4686	4687
	SEQ NO.		1172	1173	1174	1175	1176	1177	1178	1179	1180	1181	1182	1183	1184	1185	1186	1187

5
10
15
20
25
30
35
40
45
50

_		<del>,</del>				<del></del> -		- T						_	Т	Т		<del>r</del>	$\neg$
	Function	cysteine desulphurase	nicotinate-nucleotide pyrophosphorylase	quinolinate synthetase A	DNA hydrolase	hypothetical membrane protein	hypothetical protein	hypothetical protein	lipoate-protein ligase A	alkylphosphonate uptake protein and C-P lyase activity	transmembrane transport protein or 4-hydroxybenzoate transporter	p-hydroxybenzoate hydroxylase (4- hydroxybenzoate 3- monooxygenase)	hypothetical membrane protein	ABC transporter ATP-binding protein	hypothetical membrane protein	_	Ca2+/H+ antiporter ChaA	hypothetical protein	hypothetical membrane protein
	Matched length (a.a.)	376	283	361	235	192	214	108	216	148	420	395	191	532	250		339	236	221
	Similarity (%)	73.4	689	77.6	6.09	54.7	66.4	74.1	60.7	8.09	64.3	68.6	9.69	47.6	61.6		0.69	57.6	61.1
	Identity (%)	43.9	42.1	49.3	37.0	23.4	36.0	41.7	30.1	29.7	28.8	40.8	36.7	24.8	25.6		33.3	28.4	27.6
Table 1 (continued)	Homologous gene	Ruminococcus flavefaciens cysteine desulphurase gene	Mycobacterium tuberculosis	Bacillus subtilis nadA	Streptomyces coelicolor SC5B8.07	Deinococcus radiodurans R1 DR1112	Streptomyces coelicolor SC3A7.08	Escherichia coli K12 MG1655 ybdF	Escherichia coli K12 lplA	Escherichia coli K12 phnB	Pseudomonas putida pcaK	Pseudomonas aeruginosa phhy	Bacillus subtilis 168 ykoE	Escherichia coli yijK	Bacillus subtilis 168 ykoC		Escherichia coli chaA	Pyrococcus abyssi Orsay PAB1341	Baciltus subtitis ywaF
	db Match	gp:RFAJ3152_2	sp:NADC_MYCTU	pir.E69663	gp:SC5B8_7	gp:AE001961_5	gp:SC3A7_8	sp:YBDF_ECOLI	gp:AAA21740_1	sp:PHNB_ECOL!	sp:PCAK_PSEPU	sp:PHHY_PSEAE	pir.A69859	Sp:YJJK ECOLI	pir.G69858		sp:CHAA_ECOLI	pir.C75001	sp:YWAF_BACSU
	ORF (bp)	1074	837	1182	642	909	909	342	789	411	1293	1185	588	1338	753	531	1050	708	723
	Terminal (nt)	1115832	1116908	1117751	1119086	1120804	1120833	1121468	1121818	1123461	1123534	1124836	1127009	1128350	1129102	1129632	1130704	1131428	1131401
	Initial (nt)	1116905	1117744	1118932	1119727	1120205	1121432	1121809	1122606	1	1124826	1126020	1126422	1127013		1129102	1129655	1130721	1132123
	SEQ NO.		4689	4690		4692	4693	4694	4695	4696	4697	4698	4699	4700	4701	4702	4703	4704	4705
	NO NO NO NA	<del></del> -	1189	190	+ -	1192	1193	1194	1195	-	1197	1198	1199	1200	1201	1202	1203	1204	1205

,									-				E						T		$\neg$
5	Function	excinuclease ABC subunit A	thioredoxin peroxidase			hypothetical membrane protein	oxidoreductase or thiamin biosynthesis protein					chymotrypsin BII	arsenate reductase (arsenical pump modifier)	hypothetical membrane protein	hypothetical protein	hypothetical protein	GTP-binding protein (tyrosine phsphorylated protein A)	hypothetical protein	hypothetical protein		ferredoxin [4Fe-4S]
15	Matched length (a.a.)	946	164		$\neg$	318	282					27.1	111	340	147	221	614	506	315		103
20	Similarity (%)	58.7	81.7			72.0	49.0					51.3	72.1	62.4	71.4	62.9	76.7	54.9	61.9		91.3
	Identity (%)	35.5	57.3			39.9	34.0					28.8	43.2	23.5	43.5	35.8	46.3	27.9	38.7		78.6
52 Table 1 (continued)	us gene	hilus unrA	berculosis			Tpa	licolor A3(2)	-				ei		/аD	Iberculosis	Iberculosis	(12 typA	ıberculosis	ıberculosis		seus fer
30 Table 1 (0	Homologous gene	Thermus thermophilus unrA	Mycobacterium tuberculosis H37Rv tpx			Escherichia coli yedL	Streptomyces coelicolor A3(2)					Penaeus vannamei	Escherichia coli	Bacillus subtilis yyaD	Mycobacterium tuberculosis H37Rv Rv1632c	Mycobacterium tuberculosis H37Rv Rv1157c	Escherichia coli K12 typA	Mycobacterium tuberculosis H37Rv Rv1166	Mycobacterium tuberculosis H37Rv Rv1170		Streptomyces griseus fer
<i>40</i>	db Match	p:UVRA_THETH 1	sp:TPX_MYCTU			sp. YEDI_ECOLI E	gp:SCF76_2					sp.CTR2_PENVA F	sp:ARC2_ECOLI	sp.YYAD_BACSU		pir.F70555	Sp.TYPA_ECOLI	pir.F70874	pir.B70875		sp.FER_STRGR
	ORF (bp)	2340 sp	495 sp	216	1776	954 sp	900	366	297	261	387	834 sp	345 sp	1200 sp	537 pi	714 pi	1911 sp	1506 pi	870 pi	438	315 sp
45	Terminal (nt)	1132133	1135055	1135691	1135058	1136938	1138859	1139245	1139492	1139617	1139635	1140028	1140901	1142472	1142479	1143026	1146028	1147602	1148461	1148882	1149267
50	Initial (nt)	1134472	1134561	1135476	1136833	1137891	1137960	1138880	1139196	1139357	1140021	1140861	1141245	1141273	1143015	1143739	1144118	1146097	1147592	1148445	1148953
	SEQ NO.	4706	4707	4708	4709	4710	4711	4712	4713	4714	4715	4716	4717	4718	4719	4720	4721	4722	4723	4724	4725
55	SEQ NO.	1206	1207	1208	1209	1210	1211	1212	1213	1214	1215	1216	1217	1218	1219	1220	1221	1222	1223	1224	1225

EP 1 108 790 A2

5	Function	aspartate aminotransferase			tetrahydrodipicolinate succinylase or succinylase or succinylation of piperidine-2,6-dicarboxylate		hypothetical protein	dihydropteroate synthase	hypothetical protein	hypothetical protein	antigen TbAAMK, useful in vaccines for prevention or treatment of tuberculosis	mycinamicin-resistance gene	sucrose-6-phosphate hydrolase	ADPglucose-starch(bacterial glycogen) glucosyltransferase	glucose-1-phosphate adenylyltransferase	methyltransferase	RNA polymerase sigma factor (sigma-24); heat shock and oxidative stress	
15	Matched length (a.a.)	397			229		211	273	245	66	47	286	524	433	400	93	194	
20	Similarity (%)	52.9			100.0		100.0	69.0	73.1	67.7	91.5	67.8	51.0	51.3	81.8	62.4	57.2	
	Identity (%)	25.9			100.0		100.0	59.0	45.7	31.3	72.3	39.2	23.5	24.7	61.0	25.8	27.3	
os 52 Zable 1 (continued)	is gene	strain YM-2 aat			glutamicum		glutamicum	licolor A3(2)	prae u17561	berculosis	berculosis	griseorubida	osaceus scrB	(12 MG1655	elicolor A3(2)	carofaciens	poE	
e Table 1 (c	Homologous gene	Bacillus sp. strain			Corynebacterium glutamicum ATCC 13032 dapD		Corynebacterium glutamicum ATCC 13032 orf2	Streptomyces coelicolor A3(2) dhpS	Mycobacterium teprae u17561	Mycobacterium tuberculosis H37Rv Rv1209	Mycobacterium tuberculosis	Micromonospora griseorubida myrA	Pediococcus pentosaceus scrB	Escherichia coli K12 MG1655 glgA	Streptomyces coelicolor A3(2) glgC	Streptomyces mycarofaciens MdmC	Escherichia coli rpoE	
35							υ¥	00	14				$\vdash$		<del>                                     </del>	<del></del>		
40	db Match	sp:AAT_BACSP			gp:CGAJ4934_1		pir:S60064	gp:SCP8_4	qp:MLU15180	pir.G70609	gsp:W32443	sp:MYRA_MICGR	SP:SCRB_PEDPE		sp.GLGC_STRCO	sp:MDMC_STRMY	sp:RPOE_ECOLI	
	ORF (bp)	1101	621	1185	891	663	768	831	729	306	165	864	1494	<del></del>	1215	639	639	492
45	Terminal (nt)	1150379	1151028	1152370	1152373 L	1155875	1157669	1158524	1159252	1159572	1159799	1160728	1160738	1162379	1164916	1164974	1166384	1167067
50	Initial (nt)	1149279	1150408	1151186	1153263	1156537	1156902	1157694	1158524	1159267	1159635	1159865	1162231	1163605	1163702	1165612	1165746	1166576
•	SEQ NO.	4726	4727	4728	4729	4730	4731	4732	4733	4734	4735	4736	4737		4739	4740	4741	4742
55	SEO NO.	1226	1227	1228	1229	1230	1231	1232	1233	1234	1235	1236	1237	1238	1239	1240	1241	1242

EP 1 108 790 A2

5

	Function	hypothetical protein	ATPase	hypothetical protein	hypothetical protein	hypothetical protein			2-oxoglutarate dehydrogenase	ABC transporter or multidrug resistance protein 2 (P-glycoprotein 2)	hypothetical protein	shikimate dehydrogenase	para-nitrobenzyl esterase				tetracycline resistance protein	metabolite export pump of tetracenomycin C resistance	
	Matched length (a.a.)	112	257	154	434	140			1257	1288	240	255	501				409	444	
	Similarity (%)	73.2	72.0	83.8	0.77	87.1			8.66	60.4	72.1	61.2	64.7				61.4	64.2	
	Identity (%)	45.5	43.6	60.4	49.8	57.9			99.4	28.8	31.7	25.5	35.7				27.1	32.4	
Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv1224	Escherichia coli mrp	Mycobacterium tuberculosis H37Rv Rv1231c	Mycobacterium tuberculosis H37Rv Rv1232c	Mycobacterium tuberculosis H37Rv Rv1234			Corynebacterium glutamicum AJ12036 odhA	Cricetulus griseus (Chinese hamster) MDR2	Mycobacterium tuberculosis H37Rv Rv1249c	Escherichia coli aroE	Bacillus subtilis pnbA				Escherichia coli transposon Tn1721 tetA	Streptomyces glaucescens tcmA	
	db Match	pir.C70508	sp:MRP_ECOLI	pir:B70509	pir.C70509	pir.A70952			prf.2306367A	sp:MDR2_CRIGR	pir.H70953	sp:AROE_ECOLI	sp.PNBA_BACSU				sp:TCR1_ECOLI	sp.TCMA_STRGA	
	ORF (bp)	468	1125	579	1290	516	999	594	3771	3741	717	804	1611	651	876	525	1215	1347	705
	Terminal (nt)	1167577	1167587	1168747	1169321	1171187	1171871	1171869	1172501	1176308	1.180121	1180872	1183603	1184257	1185155	1185218	1187039	1188389	1190526
	Initial (nt)	1167110	1168711	1169325	1170610	1170672	1171206	1172462	1176271	1180048	1180837	1181675	1181993	1183607	1184280	1185742	1185825	1187043	1189822
	SEQ NO.	4743	4744	4745	4746	4747	4748	4749	4750	4751	4752	4753	4754	4755	4756	4757	4758	4759	4760
	SEQ NO. (DNA)	1243	1244	1245	1246	1247	1248	1249	1250	1251	1252	1253	1254	1255	1256	1257	1258	1259	1260

5		Function	5- methyltetrahydropteroyltriglutamate- -homocysteine S-methyltransferase		thiophene biotransformation protein						ABC transporter	ABC transporter	cytochrome bd-type menaquinol oxidase subunit II	cytochrome bd-type menaquinol oxidase subunit I	e v		mutator mutT protein ((7,8-dihydro-8-oxoguanine-triphosphatase)(8-oxo-dGTPase)(dGTP		proline-specific permease
15		th (-													2 helicase				$\neg$
, ,		Matched length (a.a.)	774		444						526	551	333	512	402		86		433
20		Similarity (%)	72.2		79.5						63.5	58.4	93.0	99.0	55.0		65.6		85.0
		Identity (%)	45.2		55.2						28.7	29.4	92.0	9.66	26.4		36.9		51.3
25	ontinued)	e deue	us metE		s strain KGB1						12 MG1655	12 MG1655	jlutamicum ctofermentum)	jlutamicum ctofermentum)	12 MG1655		nutT		urium proY
30	Table 1 (continued)	Homologous gene	Catharanthus roseus metE		Nocardia asteroides strain KGB1						Escherichia coli K12 MG1655 cydC	Escherichia coli K12 MG1655 cydD	Corynebacterium glutamicum (Brevibacterium lactofermentum) cydB	Corynebacterium glutamicum (Brevibacterium lactofermentum) cydA	Escherichia coli K12 MG1655 yejH		Proteus vulgaris mutT		Salmonella typhimurium proY
<i>35</i>		db Match	pir.S57636		gsp:Y29930						sp.CYDC_ECOLI	sp:CYDD_ECOLI	p:AB035086_2	Ip:AB035086_1	sp:YEJH_ECOLI		sp:MUTT_PROVU		sp:PROY_SALTY
	,	ORF (bp)	2235 p	456	1398 g	324	945	792	1647	192	1554 8	1533	666 666	1539 g	2265	342	393 8	765	1404
45		Terminal (nt)	1188388	1191542	1193807	1194190	1195109	1195125	1197620	1197815	1197990	1199543	1201090	1202094	1203916	1206657	1206831	1208138	1208212
50		Initial (nt)	1190622	1191087	1192410	1193867	1194165	1195916	1195974	1197624	1199543	1201075	1202088	1203632	1206180	1206316		1207374	1209615
		SEQ NO (a.a.)	<del></del>	4762	4763	4764	4765	4766	4767	4768	4769	4770	4771	4772	4773	4774		4776	
55		SEQ.	1261	1262	1263	1264	1265	1266	1267	1268	1269	1270	1271	1272	1273	1274	1275	1276	1277

EP 1 108 790 A2

	ſ	т				$\neg$		Т		$\neg \neg$	Т	·T			Т	1	Т		$\neg$
5		Function	DEAD box ATP-dependent RNA helicase	bacterial regulatory protein, tetR family	pentachlorophenol 4- monooxygenase	maleylacetate reductase	catechol 1,2-dioxygenase		hypothetical protein	transcriptional regulator		hypothetical protein	phosphoesterase	hypothetical protein			esterase or lipase		
15		Matched length (a.a.)	643	247	595	354	278		185	878		203	395	915			220		
20		Similarity (%)	74.3	47.4	47.7	72.0	59.4		58.4	55.4		56.2	67.3	59.6			64.6		
		Identity (%)	48.1	24.7	24.5	40.4	30.6		31.9	24.9		29.6	39.2	29.7			37.3		
<b>25</b>	Table 1 (continued)	is gene	niae CG43 spendent RNA	orae	ча рсрВ	B13 clcE	oaceticus		berculosis	erevisiae		licolor A3(2)	berculosis	berculosis			ing bacterium		
30	Table 1 (c	Homologous gene	Klebsiella pneumoniae CG43 DEAD box ATP-dependent RNA helicase deaD	Mycobacterium leprae B1308_C2_181	Sphingomonas flava pcpB	Pseudomonas sp. B13 clcE	Acinetobacter calcoaceticus catA		Mycobacterium tuberculosis H37Rv Rv2972c	Saccharomyces cerevisiae SNF2		Streptomyces coelicolor A3(2) orf2	Mycobacterium tuberculosis H37Rv Rv1277	Mycobacterium tuberculosis H37Rv Rv1278			Petroleum-degrading bacterium HD-1 hde		
35		db Match	sp:DEAD_KLEPN	prf:2323363BT	sp.PCPB_FLAS3	sp:CLCE_PSESB				p:SNF2_YEAST		p:SCO007731_6		sp:Y084_MYCTU			gp:AB029896_1		
40						-			pir.A70672	S	10	6	3 pir:E70755						
		ORF (pp)	2196	687	1590	1068	885	471	540	3102	1065	828	1173	2628	306	318	774	378	786
45		Terminal (nt)	1212129	1212429	1214858	1215938	1216836	1216904	1217443	1222996	1221841	1223843	1225059	1227693	1227282	1227340	1228636	1229095	1229935
50		Initial (nt)	1209934	1213115	1213269	1214871	1215952	1217374	1217982	1219895	1222905	1222986	1223887	1225066	1227587	1227657	1227863	1228718	1229150
		SEQ NO.	4778	4779	4780	4781	4782	4783	4784	4785	4786	4787	4788	4789	4790	4791	4792	4793	4794
55		SEQ NO. (DNA)	1278	1279	1280	1281	1282	1283	1284	1285	1286	1287	1288	1289	1290	1291	1292	1293	1294

EP 1 108 790 A2

																	_		-	-			$\neg$
5		Function	short-chain fatty acids transporter	regulatory protein			fumarate (and nitrate) reduction regulatory protein	mercuric transort protein periplasmic component precursor	zinc-transporting ATPase Zn(II)- translocating P-type ATPase	GTP pyrophosphokinase (ATP:GTP 3-pyrophosphotransferase) (ppGpp synthetase I)	tripeptidyl aminopeptidase			homoserine denydrogenase		ricely commended to	nitrate reductase gamma criain	nitrate reductase della challi	nitrate reductase beta chain	hypothetical protein	hypothetical protein	nitrate reductase alpha chain	nitrate extrusion protein
			short	regu	$\downarrow$		fuma	com	zinc- trans	3'-py	tripe	_	$\perp$	E P			nitre	I I	nitic	d y	P P	nitr	ojt.
15		Matched length (a.a.)	122	166			228	81	605	137	601			54		18	220	175	505	137	83	1271	461
20		Similarity (%)	69.7	56.6			57.9	66.7	70.6	58.4	49.3			98.0			69.6	63.4	83.4	48.0	55.0	73.8	67.9
		Identity (%)	37.7	24.7			25.0	33.3	38.0	32.9	26.6			95.0		1	45.0	30.3	56.6	36.0	36.0	46.9	32.8
25	inued)	ene	or	recS			AG1655 fnr	ens merP	MG1655		s tap			amicum						APE1291	APE1289		narK
<b>30</b>	Table 1 (continued)	Homologous gene	Streptomyces coelicolor SC1C2.14c atoE	Erwinia chrysanthemi recS			Escherichia coli K12 MG1655 fnr	Shewanella putrefaciens merP	Escherichia coli K12 MG1655 atzN	Vibrio sp. S14 relA	Streptomyces lividans tap			Corynebacterium glutamicum			Bacillus subtilis narl	Bacillus subtilis narJ	Bacillus subtilis narH	Aeropyrum pernix K1 APE1291	Aeropyrum pernix K1 APE1289	Bacillus subtilis narG	Escherichia coli K12 narK
<i>35</i>		db Match	sp:ATOE_ECOLI	p:PECS_ERWCH			p:FNR_ECOLI	p:MERP_SHEPU	p.ATZN_ECOLI	sp.RELA_VIBSS	gsp:R80504			GSP:P61449			sp:NARI_BACSU	sp:NARJ_BACSU	SP:NARH_BACSU	PIR:D72603	PIR:B72603	sp:NARG_BACSU	SP. NARK_ECOL!
		ORF (bp)	537 sp.	486 sp	222	519	750 sp	234 sp	1875 sp	e30 sp	1581 gs	603	120	108 G	1260	069	777 sp	732 sp	1593 sr	594 P	+	+-	1350 sp
<b>45</b>			<u> </u>	-	╁			+	<del> </del>	<del> </del>	+	+	-	$\vdash$			-	_	-	+	+	1	1
-		Terminal (nt)	1229180	1230480	1230831	1230914	1232479	1232836	1234881	1235612	1236545	1241554	1242156	1243728	1243942	1244843	1245720	1246508	+-	1250444	+		1252557
50		Initial (nt)	1229716	1229995	1230610	1231432	1231730	1232603	1233007	1234983	1238125		1242275	1243621	1245201	1245532	1246496	1247239					1253906
		SEQ NO.	+	4796	4797	4798		4800	4801	4802	4803	4804	4805	4806	4807	4808	4809	4810		4812		_	
		008		96	766	86	599	8	39.	302	303	304	305	306	307	308	309	310	311	312	313	314	315

undecaprenyl-phosphate alpha-N-acetylglucosaminyltransferase

322

58.4

31.1

Escherichia coli K12 rfe

Sp.RFE\_ECOLI

1146

1271192

1270047

4833

1333

hypothetical protein

215

86.0

62.3

Mycobacterium tuberculosis H37Rv Rv1301

sp:YD01\_MYCTU

648

1270043

1269396

4832

1332

molybdopterin guanine dinucleotide synthase edium-chain fatty acid-CoA ligase molybdopterin biosynthsisi protein Moybdenume (mosybdenum molybdoptein biosynthesis protein molybdopterin biosynthesis cnx1 hypothetical membrane protein hypothetical membrane protein protein (molybdenum cofactor peptide chain release factor 1 cofastor biosythesis enzyme) extracellular serine protease protoporphyrinogen oxidase 5 biosynthesis enzyme cnx1) Function 10 precurosor Rho factor Matched length (a.a.) 15 572 363 280 472 753 738 334 366 354 157 Similarity 73.8 71.9 45.9 58.2 o, 62.6 65.7 65.0 73.7 8 52. 60 57. 20 Identity (%) 41.9 32.5 31.6 27.5 32.8 51.4 36.7 50.7 31.1 21.1 œ 8 Serratia marcescens strain IFO-3046 prtS 25 Arabidopsis thaliana CV cnx1 Mycobacterium tuberculosis H37Rv Rv1842c Table 1 (continued) Mycobacterium tuberculosis H37Rv Rv1841c Mycobacterium tuberculosis H37Rv Rv0438c moeA Pseudomonas putida mobA Pseudomonas ofeovorans Escherichia coli K12 RF-1 Arabidopsis thaliana cnx2 Homologous gene Micrococcus luteus rho Escherichia coli K12 30 35 sp:CNX2\_ARATH sp:PRTS\_SERMA Sp:CNX1\_ARATH sp:Y0D3\_MYCTU gp:PPU242952\_2 sp:Y0D2\_MYCTU sp:MOEA\_ECOLI SP.ALKK\_PSEOL Sp:HEMK\_ECOLI Sp.RHO\_MICLU sp:RF1\_ECOLI db Match 40 1131 1725 2286 1074 1008 1209 1866 1401 1023 837 489 969 774 ORF (bp) 561 603 684 1268267 1259429 1262886 1268503 1269343 1257750 1257865 1259993 1261688 1267427 1265611 1265427 1256851 1266267 1254634 1254737 Terminal 45 £ 1262818 1264610 1266449 1269040 1259265 1259989 1265142 1266306 1267430 1254146 1256602 1257067 1257858 1261201 1265665 1268507 Initial (nt) 50 4823 4825 4828 4829 4830 4816 4819 4824 4826 4827 4817 4818 4820 4822 4831 4821 SEQ NO. 1325 1326 1324 1327 1328 1329 1330 (DNA) 1316 1317 1318 1319 1320 1321 1322 1323 1331 SEQ 9

												· · ·						
5		Function		hypothetical protein	ATP synthase chain a (protein 6)	H+-transporting ATP synthase lipid- binding protein. ATP synthase C chane	H+-transporting ATP synthase chain b	H+-transporting ATP synthase delta chain	H+-transporting ATP synthase alpha chain	H+-transporting ATP synthase gamma chain	H+-transporting ATP synthase beta chain	H+-transporting ATP synthase epsilon chain	hypothetical protein	hypothetical protein	putative ATP/GTP-binding protein	hypothetical protein	hypothetical protein	thioredoxin
15		Matched length (a.a.)		80 hy	245 A1	7	151 H	274 Ch	516 H	320 H	483 Ch	122 H-	132 hy	230 hy	95 pr	134 hy	101 hy	301 th
20		Similarity (%)		99.0	26.7	85.9	6.99	67.2	88.4	9.92	100.0	73.0	67.4	85.7	26.0	68.7	79.2	71.4
		Identity (%)		98.0	24.1	54.9	27.8	34.3	6.99	46.3	99.8	41.0	38.6	70.0	45.0	35.8	54.5	37.9
25	tinued)	jene		tamicum	atpB	s atpL	s atpF	s atpD	s atpA	s atpG	tamicum	s atpE	culosis	culosis	olor A3(2)		culosis	culosis
30	Table 1 (continued)	Homologous gene		Corynebacterium glutamicum atpl	Escherichia coli K12 atpB	Streptomyces lividans atpL	Streptomyces lividans atpF	Streptomyces lividans atpD	Streptomyces lividans atpA	Streptomyces lividans atpG	Corynebacterium glutamicum AS019 atpB	Streptomyces lividans atpE	Mycobacterium tuberculosis H37Rv Rv1312	Mycobacterium tuberculosis H37Rv Rv1321	Streptomyces coelicolor A3(2)	Bacillus subtilis yqjC	Mycobacterium tuberculosis H37Rv Rv1898	Mycobacterium tuberculosis H37Rv Rv1324
35		db Match		GPU:AB046112_1	p:ATP6_ECOU		p:ATPF_STRLI	p:ATPD_STRLI	STRLI	SP.ATPG_STRLI	p:ATPB_CORGL	SP.ATPE_STRLI	sp:Y0ZW_MYCTU	sp:Y036_MYCTU	GP.SC26G5_35	sp:YQJC_BACSU	sp.YC20_MYCTU	sp:YD24_MYCTU
40				<del>-</del>	5	<u> </u>	, v	- 0,	4 sp:ATPA_8		ļ.,	-	<del></del>		Ť	-		
		ORF (bp)	486	249	810	240	564	813	1674	975	1449	372	471	069	285	453	312	921
45		Terminal (nt)	1271698	1272119	1273149	1273525	1274122	1274943	1276648	1277682	1279136	1279522	1280240	1280959	1281251	1281262	1282105	1283114
50		Initial (nt)	1271213	1271871	1272340	1273286	1273559	1274131	1274975	1276708	1277688	1279151	1279770	1280270	1280967	1281714	1281794	1282194
		SEQ NO.	4834	4835	4836	4837	4838	4839	4840	4841	4842	4843	4844	4845	4846	4847	4848	4849
55		SEQ NO.	1334	1335	1336	1337	1338	1339	1340	1341	1342	1343	1344	1345	1346	1347	1348	1349

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

nued)
(contri
Table 1

	Function	FMNH2-dependent aliphatic sulfonate monooxygenase	alphatic sulfonates transport permease protein	alphatic sulfonates transport permease protein	sulfonate binding protein precursor	1,4-alpha-glucan branching enzyme (glycogen branching enzyme)	alpha-amylase	•	ferric enterobactin transport ATP- binding protein or ABC transport ATP-binding protein	hypothetical protein	hypothetical protein		electron transfer flavoprotein beta- subunit	electron transfer flavoprotein alpha subunit for various dehydrogenases		nitrogenase cofactor sythesis protein		hypothetical protein	
	Matched length (a.a.)	366	240	228	311	710	467		211	260	367		244	335		375		397	
	Similarity (%)	74.3	75.8	72.8	62.1	72.7	50.5		87.6	68.5	70.0		64.8	61.8		67.7		55.7	
	Identity (%)	50.3	40.8	50.4	35.1	46.1	22.9		31.8	39.6	43.1		31.2	33.1		35.2		29.5	
(100)	Homologous gene	Escherichia coli K12 ssuD	Escherichia coli K12 ssuC	Escherichia coli K12 ssuB	Escherichia coli K12 ssuA	Mycobacterium tuberculosis H37Rv Rv1326c glgB	Dictyoglomus thermophilum amyC		Escherichia coli K12 fepC	Mycobacterium tuberculosis H37Rv Rv3040c	Mycobacterium tuberculosis H37Rv Rv3037c		Rhizobium meliloti fixA	Rhizobium meliloti fixB		Azotobacter vinetandii nifS		Rhizobium sp. NGR234 plasmid pNGR234a y4mE	
	db Match	gp:ECO237695_3	sp:SSUC_ECOLI	sp:SSUB_ECOLI	SP.SSUA ECOLI	sp:GLGB_ECOLI	sp:AMY3_DICTH		sp:FEPC_ECOLI	pir.C70860	pir:H70859		sp:FIXA_RHIME	sp.FIXB_RHIME		sp:NIFS_AZOVI		sp:Y4ME_RHISN	
	ORF (bp)	1143	768	729	957	2193	1494	348	879	804	1056	612	786	951	615	1128	312	1146	١.
	Terminal (nt)	1284466	1285284	1286030	1286999	1287281	1289514	1291373	1292577	1294025	1295206	1294436	1296220	1297203	1297093	1298339	1298342	1299000	
	Initial (nt)	1283324	1284517	1285302	1286043		1291007	1291026		1293222	1294151	1295047	1295435	1296253	1296479	1297212	1298653	1300145	
	SEQ NO.	4850	4851	4852	4853	4854	4855	4856	4857	4858	4859	4860	4861	4862	4863	4864	4865	4866	
	SEO NO.		1351	1352	1353		1355	1356		1358	1359	1360	1361	1362	1363	1364	1365	1366	-

5
10
15
20
25
30
35
40
45
50

-																		
	Function	transcriptional regulator	acetyltransferase				tRNA (5-methylaminomethyl-2- thiouridylate)-methyltransferase		hypothetical protein	tetracenomycin C resistance and export protin		DNA ligase (polydeoxyribonucleotide synthase [NAD+]	hypothetical protein	glutamyl-tRNA(Gln) amidotransferase subunit C	glutamyl-tRNA(Gln) amidotransferase subunit A	vibriobactin utilization protein / iron- chelator utilization protein	hypothetical membrane protein	pyrophosphatefructose 6- phosphate 1-phosphotransrefase
	Matched length (a.a.)	59	181				361		332	200		677	220	65	484	263	96	358
	Similarity (%)	76.3	55.3				80.9		66.0	65.8		70.6	70.9	64.0	83.0	54.0	79.2	77.9
	Identity (%)	47.5	34.8				61.8		33.7	30.2		42.8	40.0	53.0	74.0	28.1	46.9	54.8
Table 1 (continued)	Homologous gene	Rhizobium sp. NGR234 plasmid pNGR234a Y4mF	Escherichia coli K12 MG1655 yhbS				Mycobacterium tuberculosis H37Rv Rv3024c		Mycobacterium tuberculosis H37Rv Rv3015c	Streptomyces glaucescens tcmA		Rhodothermus marinus dnlJ	Mycobacterium tuberculosis H37Rv Rv3013	Streptomyces coelicolor A3(2) gatC	Mycobacterium tuberculosis H37Rv gatA	Vibrio vulnificus viuB	Streptomyces coelicolor A3(2) SCE6.24	Amycolatopsis methanolica pfp
	db Match	sp:Y4MF_RHISN	sp:YHBS_ECOLI				pir.C70858		pir:870857	sp:TCMA_STRGA		sp:DNLJ_RHOMR	pir.H70856	sp:GATC_STRCO	sp.GATA_MYCTU	sp:VIUB_VIBVU	gp:SCE6_24	sp:PFP_AMYME
	ORF (bp)	225	504	942	1149	396	1095	654	066	1461	735	2040	663	297	1491	849	306	1071
	Terminal (nt)	1300145	1301055	1300988	1301975	1303694	1304923	1303883	1305921	1305924	1307462	1310369	1310435	1311616	1313115	1314118	1314470	1316083
	Initial (nt)	1300369	1300552	1301929	1303123	1303299	1303829	1304536	1304932	1307384	1308196	1308330	1311097	1311320	1311625	1313270	1314775	1315013
	SEQ NO.	4867	4868	4869	4870	4871	4872	4873	4874	4875	4876	4877	4878	4879	4880	4881	4882	4883
	SEQ NO.	1367	1368	1369	1370	1371	1372	1373	1374	1375	1376		1378	1379	1380	1381	1382	1383

											.—								
5		Function		glucose-resistance amylase regulator (catabolite control protein)	ripose transport ATP-binding protein	high affinity ribose transport protein	periplasmic ribose-binding protein	high affinity ribose transport protein	hypothetical protein	iron-siderophore binding lipoprotein	Na-dependent bile acid transporter	RNA-dependent amidotransferase B	putative F420-dependent NADH reductase	hypothetical protein	hypothetical protein	hypothetical membrane protein		dihydroxy-acid dehydratase	hypothetical protein
15		Matched length (a.a.)		328 9	499 ri	329 h	305 p	139 h	200 h	354 ir	268 N	485 R	172 P	317 h	234 h	325 h		613 d	105 h
20		Similarity (%)		31.4	76.2	76.9	1.77	68.4	58.0	60.2	61.9	71.8	61.1	6.99	62.4	52.6		99.4	68.6
		Identity (%)		31.4	44.7	45.6	45.9	41.7	31.0	31.4	35.8	43.1	32.6	39.8	39.3	27.4		99.2	33.3
25	ontinued)			п ссрА	2 rbsA	2 MG1655	2 MG1655	2 MG1655	evisiae	color	Rat) NTCI	eus WHU 29	naschii	2 yajG	erculosis	erculosis		utamicum	erculosis
30	Table 1 (continued)	Homologous gene	,	Bacillus megaterium ccpA	Escherichia coli K12 rbsA	Escherichia coli K12 MG1655 rbsC	Escherichia coli K12 MG1655 rbsB	Escherichia coli K12 MG1655 rbsD	Saccharomyces cerevisiae YIR042c	Streptomyces coelicolor SCF34,13c	Rattus norvegicus (Rat) NTCI	Staphylococcus aureus WHU 29 ratB	Methanococcus jannaschii MJ1501 f4re	Escherichia coli K12 yajG	Mycobacterium tuberculosis H37Rv Rv2972c	Mycobacterium tuberculosis H37Rv Rv3005c		Corynebacterium glutamicum ATCC 13032 ilvD	Mycobacterium tuberculosis H37Rv Rv3004
35					<u> </u>			İ		တတ	2	N 15			ΣI	ΣI			ΣI
40		db Match		sp.CCPA_BACME	sp.RBSA_ECOLI	sp:RBSC_ECOLI	sp:RBSB_ECOLI	sp:RBSD_ECOLI	sp:YIW2_YEAST	gp:SCF34_13	sp:NTCI_RAT	gsp:W61467	sp:F4RE_METJA	sp:YaJG_ECOLI	pir.A70672	pir.H70855		gp:AJ012293_1	pir.G70855
		ORF (bp)	630	1107	1572	972	942	369	636	1014	1005	1479	672	1077	774	1056	237	1839	564
45		Terminal (nt)	1315325	1317444	1319005	1319976	1320942	1321320	1322111	1323406	1324537	1326256	1327049	1329891	1331875	1333008	1333188	1333442	1335412
50		Initial (nt)	1315954	1316338	1317434	1319005	1320001	1320952	1321476	1322393	1323533	1324778	1326378	1330967	1331102	1331953	1333424	1335280	1335975
		SEQ NO. (a.a.)	4884	4885	4886	4887	4888	4889	4890	4891	4892	4893	4894	4895	4896	4897	4898	4899	4900
55		SEQ NO. (DNA)	1384	1385	1386	1387	1388	1389	1390	1391	1392	1393	1394	1395	1396	1397	1398	1399	1400

5	
10	
15	
20	
25	
<b>30</b>	
35	
40	
<b>45</b> .	
50	

	Function	hypothetical membrane protein	hypothetical protein		nitrate transport ATP-binding potein	maltose/maltodextrin transport ATP- binding protein	nitrate transporter protein			actinorhodin polyketide dimerase	cobalt-zinc-cadimium resistance protein			hypothetical protein		D-3-phosphoglycerate dehydrogenase	hypothetical serine-rich protein			hypothetical protein	
	Matched length (a.a.)	62	99		167	87	324			142	304			642		530	105			620	
	Similarity (%)	100.0	55.0		80.8	78.2	56.8			73.2	72.7			53.7		100.0	52.0			63.1	
	Identity (%)	100.0	45.0		50.9	46.0	28.1			39.4	39.1			22.9		93.8	29.0			32.9	
Table 1 (continued)	Homologous gene	Corynebacterium glutamicum ATCC 13032 yilV	Sulfolobus solfataricus		Synechococcus sp. nrtD	Enterobacter aerogenes (Aerobacter aerogenes) malK	Anabaena sp. strain PCC 7120 nrtA			Streptomyces coelicolor	Ralstonia eutropha czcD			Methanococcus jannaschii		Brevibacterium flavum serA	Schizosaccharomyces pombe SPAC11G7.01			Rhodobacter capsulatus strain SB1003	
	db Match	sp:YILV_CORGL	GP:SSU18930_26 3	·	sp:NRTD_SYNP7	sp:MALK_ENTAE	sp:NRTA_ANASP			sp:DIM6_STRCO	sp:CZCD_ALCEU			sp:Y686_METJA		gsp:Y22646	SP:YEN1_SCHPO			pir.T03476	
	ORF (bp)	1473	231	909	498	267	882	447	369	486	954	153	069	1815	1743	1590	327	867	1062	1866	402
	Terminal (nt)	1336095	1338379	1342677	1341960	1342461	1342794	1344464	1344808	1345420	1346439	1345335	1345642	1348272	1350076	1352444	1351727	1353451	1354540	1357554	1356853
	Initial (nt)	1337567	1338609	1342072	1342457	1342727	1343675	1344018	1344440	1344935	1345486	1345487	1346331	1346458	1348334	1350855	1352053	1352585	1355601	1355689	1356452
	SEQ NO.	4901	4902	4903	4904	4905	4906	4907	4908	4909	4910	4911	4912	4913	4914	4915	4916	4917	4918	4919	4920
	SEQ NO. (DNA)	1401	1402	1403	1404	1405	1406	1407	1408	1409	1410	1411	1412	1413	1414	1415	1416	1417	1418	1419	1420

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

	Function		homoprotocatechiuate catabolism bifunctional isomerase/decarboxylase [includes: 2-hydroxyhepta-2, 4-diene-1, 7-dioate isomerase(hhdd isomerase); 5-carboxymethyl-2-oxo-hex-3-ene-1, 7-dioate decarboxylase(opet decarboxylase)	methyltransferase or 3- demethylubiquinone-9 3-O- methyltransferase	isochorismate synthase	glutamyl-tRNA synthetase	transcriptional regulator													thiamin biosynthesis protein
	Matched length (a.a.)		228	192	371	485	29													299
	Similarity (%)		59.2	55.7	70.4	2.69	90.0													81.0
	Identity (%)		33.3	23.4	38.0	37.3	77.0													65.1
Table 1 (continued)	Homologous gene		Escherichia coli C hpcE	Escherichia coli K12	Bacillus subtilis dhbC	Bacillus subtilis gftX	Streptomyces coelicolor A3(2)													Bacillus subtilis thiA or thiC
	db Match		sp:HPCE_ECOL!	sp:UBIG_ECOLI	Sp:DHBC BACSU	Sp.SYE BACSU	gp:SCJ33_10													sp:THIC_BACSU
	ORF (bp)	654	804	618	1128	1488	213	516	522	342	621	303	180	330	213	183	318	1152	324	1761
	Terminal (nt)	1358210	1359062	1359669	1360168	1362848	1362926	1363142	1363732	1365256	1364340	1364878	1365217	1366137	1367505	1367888	1368395	1369551	1369874	1369877
•	Initial (nt)	1357557	4922 1358259	1359052	1361295			1363657	1364253	1364915	1364960	1365180	1365396	1365808	1367293	1368070	1368078	1368400	1369551	1371637
	SEQ NO (a.a.)	<del>;                                    </del>	4922	4923	4924	4925	4926	4927	4928	4929	4930	4931	4932	4933	4934	4935	4936	4937	4938	4939
	SEQ NO.			1423	1424	1475	1426		_	1429	1430	1431	1432	1433	1434	1435	1436	1437	1438	1439

EP 1 108 790 A2

5	Function			lipoprotein		glycogen phosphorylase			hypothetical protein	hypothetical membrane protein		guanosine 3',5'-bis(diphosphate) 3'- pyrophosphatase	acetate repressor protein	3-isopropylmalate dehydratase large subunit	3-isopropylmalate dehydratase small subunit		mutator mutT protein ((7,8-dihydro-8-oxoguanine-triphosphatase)(8-oxo-dGTPase)(dGTP		NAD(P)H-dependent dihydroxyacetone phosphate reductase	D-alanine-D-alanine ligase
15	Matched length			44		797			299	256		178	257	473	195		294		331	374
20	Similarity (%)			74.0		74.0			52.8	64.8		60.1	60.7	87.5	89.2		71.4		72.2	67.4
	Identity (%)			61.0		44.2			25.4	25.4		29.8	26.1	68.1	67.7		45.9		45.0	40.4
30 For State of State	Homologous gene			Chlamydia trachomatis		Rattus norvegicus (Rat)			Bacillus subtilis yrkH	Methanococcus jannaschii Y441		Escherichia coli K12 spoT	Escherichia coli K12 iclR	Actinoplanes teichomyceticus leu2	Salmonella typhimurium		Mycobacterium tuberculosis H37Ry MLCB637.35c		Bacillus subtilis gpdA	Escherichia coli K12 MG1655 ddIA
40	db Match			GSP:Y37857		sp:PHS1_RAT			sp:YRKH_BACSU	sp:Y441_METJA		sp:SPOT_ECOLI	Sp:ICLR_ECOLI	sp:LEU2_ACTTI	sp:LEUD_SALTY		gp:MLCB637_35		sp:GPDA_BACSU	sp:DDLA_ECOLI
	ORF (bp)	348	531	132	936	2427	183	156	1407	750	477	564	705	1443	591	318	954	156	966	1080
45	Terminal (nt)	1371979	1373131	1373929	1375491	1373350	1375805	1375933	1376149	1377666	1378466	1379566	1379555	1381882	1382492	1382502	1382845	1384085	1385125	1386232
50	Initial (nt)	1372326	1372601	1373798	1374556	1375776	1375987	1376088	1377555	1378415	1378942	1379003	1380259	1380440	1381902	1382819	1383798	1383930	1384130	1385153
	SEO NO.	4940	4941	4942	4943	4944	4945	4946	4947	4948	4949	4950	4951	4952	4953	4954	4955	4956	4957	4958
55	SEQ.	(UNA)	-	1442	1443	1444	1445	1446	1447	1448	1449	1450	1451	1452	1453	1454	1455	1456	1457	1458

5	-	Function		thiamin-phosphate kinase	uracil-DNA glycosylase precursor	hypothetical protein	ATP-dependent DNA helicase	polypeptides predicted to be useful antigens for vaccines and diagnostics	biotin carboxyl carrier protein	/lase	lipopolysaccharide core biosynthesis protein		Neisserial polypeptides predicted to be useful antigens for vaccines and diagnostics	ABC transporter or glutamine ABC transporter, ATP-binding protein	nopaline transport protein	glutamine-binding protein precursor		hypothetical membrane protein		phage integrase
15		Matched length (a.a.)		335 thiam	245 uracil	568 hypot	693 ATP-0	polyp 108 antige diagn	67 biotin	167 methylase	155 lipopoly protein		Neiss 65 be us diagn	252 ABC trans	220 nopa	234 glutar		322 hypol		223 phag
20		Similarity Mi (%)		97.9	59.6	56.3	0.09	48.0	67.2	63.5	78.7		74.0	78.6	75.0	59.0		60.3		52.5
		Identity (%)		32.2	38.8	23.1	35.4	31.0	38.8	37.1	42.6		0.79	56.4	32.7	27.4		28.6		26.9
25	Table 1 (continued)	ous gene		<12 thiL	gı	italium (SGC3)	<12 recG	jitidis	n freudenreichii i	<12 yhhF	K12 MG1655		hoeae	stearothermophilus	ımefaciens	K12 MG1655		m icum MTH465		54a vinT
<i>30</i> <i>35</i>	Table 1	Homologous gene		Escherichia coli K12 thil	Mus musculus ung	Mycoplasma genitalium (SGC3) MG369	Escherichia coli K12 recG	Neisseria meningitidis	Propionibacterium freudenreichii subsp. Shermanii	Escherichia coli K12 yhhF	Escherichia coli K12 MG1655 kdtB		Neisseria gonorrhoeae	Bacillus stearoth	Agrobacterium tumefaciens nocM	Escherichia coli K12 MG1655 glnH		Methanobacterium thermoautotrophicum MTH465		Bacteriophage L54a vinT
40	,	db Match		Sp:THIL_ECOLI	sp.UNG_MOUSE	sp:Y369_MYCGE	Sp. RECG_ECOLI	GSP:Y75303	SP. BCCP_PROFR	Sp:YHHF_ECOLI	SP.KDTB_ECOLI	-	GSP:Y75358	SP.GLNQ_BACST	sp:NOCM_AGRT5	SP GLNH_ECOLI		pir:H69160		sp:VINT_BPL54
		ORF (bp)	978	993	762	1581	2121	<del></del>	213	582	480	1080		750	843	861	807	978	408	756
45		Terminal (nt)	1386293	1388324	1389073	1390788	1392916	1391638	1393151	1393735	1394221	1395933	1395097	1394800	1395568	1396561	1398468	1398557	1401333	1400185
50		Initial (nt)	1387270	1387332	1388312	1389208	1390796		1392939	1393154	1393742	1394854	1	1395549	1396410	1397421	1397662	1399534	1400926	4976 1400940
		SEQ NO.	4959	4960	4961	4962	4963	<del></del>	4965	4966	4967	4968		4970	4971	4972	4973	4974	4975	+∹
55		SEQ NO.	1459	1460	1461	1462	1463	1464	1465	1466	1467	1468	1469	1470	1471	1472	1473	1474	1475	1476

EP 1 108 790 A2

5	Function						insertion element (IS3 related)		hypothetical protein										DNA polymerase I	cephamycin export protein	DNA-binding protein	morphine-6-dehydrogenase	
15	Matched length (a.a.)						26		37										968	456	283	284	
20	Similarity (%)						96.2		97.0										80.8	67.8	65.4	76.1	
	Identity (%)						88.5		89.0										56.3	33.8	41.3	46.5	
25 (panuijuned	s gene						lutamicum		utamicum										erculosis	ndurans	color A3(2)	a morA	
30 (Continued)	Homologous gene						Corynebacterium glutamicum orf2		Corynebacterium glutamicum										Mycobacterium tuberculosis polA	Streptomyces lactamdurans cmcT	Streptomyces coelicolor A3(2) SCJ9A.15c	Pseudomonas putida morA	
40	db Match			,			pir:S60890		PIR:S60890										sp.DPO1_MYCTU	sp:CMCT_NOCLA	gp:SCJ9A_15	sp:MORA_PSEPU	
	ORF (bp)	744	432	507	864	219	192	855	111	369	315	321	375	948	306	564	222	291	2715	1422	606	873	159
45	Terminal (nt)	1402076	1402703	1402368	1403991	1404215	1404694	1405320	1406999	1407167	1407559	1408703	1409428	1410064	1411119	1411437	1412572	1412626	1416459	1416462	1418870	1419748	1419878
50	initial (nt)	1401333	1402272	1402874	1403128	1403997	1404885	1406174	1407109	1407535	1407873	1409023	1409802	1411011	1411424	1412000	1412351	1412916	1413745	1417883	1417962	1418876	1420036
	SEQ NO.	4977	4978	4979	4980	4981	4982	4983	4984	4985	4986	4987	4988	4989	4990	4991	4992	4993	4994	4995	4996	4997	4998
55	SEQ NO. (DNA)	1477	1478	1479	1480	1481	1482	1483	1484	1485	1486	1487	1488	1489	1490	1491	1492	1493	1494	1495	1496	1497	1498

								$\overline{}$				- 7		$\overline{}$					$\overline{}$			
5	Function	hypothetical protein	30S ribosomal protein S1		hypothetical protein					inosine-undine preferring nucleoside hypolase (purine nucleosidase)	aniseptic resistance protein	ribose kinase	criptic asc operon repressor, ranscription regulator		excinuclease ABC subunit B	hypothetical protein	hypothetical protein	hypothetical protein		hypothetical protein	hypothetical protein	hydrolase
15	Matched length (a.a.)	163	451		195					310	517	293	337		671	152	121	279		839	150	214
20	Similarity (%)	58.3	71.4		93.9					81.0	53.8	9'.29	65.6		83.3	59.2	80.2	77.1		47.2	68.0	58.4
	Identity (%)	31.9	39.5		80.5					61.9	23.6	35.5	30.0		57.4	33.6	38.8	53.8		23.2	32.7	30.4
Table 1 (continued)	ons gene	elicolor	K12 rpsA		actofermentum 5E					ata iunH	aureus	K12 rbsK	K12 ascG		neumoniae ) uvrB	jannaschii	K12 ytfH	K12 ytfG		yvgS	oelicolor A3(2)	K12 ycbL
	Homologous gene	Streptomyces coelicolor SCH5.13 yafE	Escherichia coli K12 rpsA		Brevibacterium lactofermentum ATCC 13869 yacE					Crithidia fasciculata iunH	Staphylococcus aureus	Escherichia coli K12 rbsK	Escherichia coli K12 ascG		Streptococcus pneumoniae plasmid pSB470 uvrB	Methanococcus jannaschii MJ0531	Escherichia coli K12 ytfH	Escherichia coli K12 ytfG		Bacillus subtilis yvgS	Streptomyces coelicolor A3(2) SC9H11.26c	Escherichia coli K12 ycbl
40	db Match	sp:YAFE_ECOLI	sp.RS1_ECOLI		sp:YACE_BRELA					sp:IUNH_CRIFA	SP. QACA STAAU	SP:RBSK_ECOLI	sp.ASCG_ECOLI		sp.UVRB_STRPN	sp:Y531_METJA	Sp:YTFH_ECOLI	sp:YTFG_ECOLI		pir:H70040	gp:SC9H11_26	sp:YCBL_ECOLI
	ORF (bp)	654 s	1458 s	1476	s 009	1098	582	246	957	936	1449	$\overline{}$	1038	798	2097	441	381	846	684	2349	912	909
45	Terminal (nt)	1420071	1422556	1421096	1425878	1427354	1427376	1427804	1429246	1428224	1429194	1430659	1431575	1433547	1436201	1436775	1436869	1438201	1440026	1438212	1440675	1441793
50	Initial (nt)	1420724	1421099	1422571	1425279	1426257	1427957	1428049	1428290	1429159	1430642		1432612	1432750	1434105	1436335	1437249	1437356	1439343	1440560	1441586	1442392
	SEQ NO.	4999	2000	5001	5005	5003	5004	5005	5006	5007	5008	_		5011	5012	5013	5014		5016	-		5019
55	SEQ NO.	1499	1500	-	1502	1503	1504	1505	1506	1507	1508	1509	1510	1511	1512	1513	1514	1515	1516	1517	1518	1519

	_											_								
5		Function	subunit A	n 1246 (uvrA	n 1246 (uvrA			n factor IF-3	tein L35	tein L20			phate transport	sphate transport	sphate transport proein	sphate transport in	Ľ	l diester e	20-}-	phenylalanyl-tRNA synthetase alpha chain
10		Func	excinuclease ABC subunit A	hypothetical protein 1246 (uvrA region)	hypothetical protein 1246 (uvrA region)			translation initiation factor IF-3	50S ribosomal protein L35	50S ribosomal protein L20			sn-glycerol-3-phosphate transport system permease protein	sn-glycerol-3-phosphate transport system protein	sn-glycerol-3-phosphate transport system permease proein	sn-glycerol-3-phosphate transport ATP-binding protein	hypothetical protein	glycerophosphoryl diester phosphodiesterase	tRNA(guanosine-2'-0-)- methlytransferase	phenylalanyl-tRN/ chain
15		Matched length (a.a.)	952	100	142			179	90	117			292	270	436	393	74	244	153	
20		Similarity (%)	9.08	67.0	47.0			78.2	76.7	92.7			71.6	70.4	57.6	71.3	56.0	20.0	71.2	
		Identity (%)	56.2	40.0	31.0			52.5	41.7	75.0			33.2	33.3	26.6	44.0	47.0	26.2	34.0	
<i>25</i>	(nanun	ene	ıvrA					ides infC	ans	ae pv.			MG1655	MG1655	MG1655	MG1655	APE0042		MG1655	syfA
30	lable I (continued)	Homologous gene	Escherichia coli K12 uvrA	Micrococcus luteus	Micrococcus luteus			Rhodobacter sphaeroides infC	Mycoplasma fermentans	Pseudomonas syringae pv. syringae			Escherichia coli K12 MG1655 ugpA	Escherichia coli K12 MG1655 upgE	Escherichia coli K12 MG1655 ugp8	Escherichia coli K12 MG1655 ugpC	Aeropyrum pernix K1 APE0042	Bacillus subtilis glpQ	Escherichia coli K12 MG1655 trmH	Bacillus subtilis 168 syfA
35	-			<u> </u>	_≥												4	<del></del>		
40		db Match	Sp.UVRA ECOLI	PIR JQ0406	PIR:JQ0406			Sp.IF3_RHOSH	SP. RL35_MYCFE	sp.RL20_PSESY			sp:UGPA_ECOLI	sp:UGPE_ECOU	sp:UGPB_ECOLI	sp:UGPC_ECOLI	PIR:E72756	sp:GLPQ_BACSU	Sp. TRMH_ECOLI	sp:SYFA_BACSU
		ORF (bp)	2847	306	450	717	2124	567	192	381	822	267	903	834	1314	1224	249	717	594	1020
45		Terminal (nt)	1445333	1443810	1444944	1446874	1445323	1448358	1448581	1449025	1449119	1450692	1451820	1452653	1454071	1455338	1454102	1455350	1456948	1458066
50		Initial (nt)	1442487	1444115	1445393	1446158	1447446	1447792	1448390	1448645	1449940	1450126	1450918	1451820	1452758	1454115	1454350	1456066	1456355	1457047
		SEQ NO.	5020		5022	5023	5024	5025	5026	5027	5028	5029	5030	5031	5032	5033	5034	5035	5036	5037
55		SEQ NO. (DNA)	1520	1521	1522	1523	1524	1525	1526	1527	1528	1529	1530	1531	1532	1533	1534	1535	1536	1537

5		Function	phenylalanyl-tRNA synthetase beta chain			cytransferase		N-acetylglutamate-5-semialdehyde dehydrogenase	tyltransferase	acetylornithine aminotransferase	te synthetase		te lyase				Kein	tyrosyl-tRNA synthase (tyrosine tRNA ligase)	stein		otein
10		Pu	phenylalanyl-tRN chain		esterase	macrolide 3-O-acytransferase		N-acetylglutama dehydrogenase	glutamate N-acetyltransferase	acetylornithine a	argininosuccinate synthetase		argininosuccinate lyase				hypothetical protein	tyrosyl-tRNA sy tRNA ligase)	hypothetical protein		hypothetical protein
15	Matched	Matched length (a.a.)	343		363	423		347	388	391	401		478				20	417	149		42
20		Similarity (%)	71.7		55.1	56.3		99.1	99.7	99.2	99.5		90.0				72.0	79.6	64.4		75.0
		Identity (%)	42.6		26.5	30.0		98.3	99.5	0.66	99.5		83.3				48.0	48.4	26.9		71.0
25 G	(2000)	Jene	MG1655		s estA	ofaciens		tamicum	tamicum	tamicum	tamicum		glutamicum				ycaR		ıaschii		m Nigg
30 tolder	lable 1 (col	Homologous gene	Escherichia coli K12 MG1655 syf8		Streptomyces scabies estA	Streptomyces mycarofaciens mdmB		Corynebacterium glutamicum ASO19 argC	Corynebacterium glutamicum ATCC 13032 argJ	Corynebacterium glutamicum ATCC 13032 argD	Corynebacterium glutamicum ASO19 argG		Corynebacterium glu ASO19 argH				Escherichia coli K12 ycaR	Bacillus subtilis syy1	Methanococcus jannaschii MJ0531		Chlamydia muridarum Nigg TC0129
40		db Match	SP. SYFB_ECOLI		SP.ESTA_STRSC	sp:MDMB_STRMY		gp:AF005242_1	sp:ARGJ_CORGL	sp:ARGD_CORGL	sp:ASSY_CORGL		gp:AF048764_1				sp:YCAR_ECOLI	sp:SYY1_BACSU	sp:Y531_METJA		PIR:F81737
		ORF (bp)	2484	77.1	972	1383	402	1041	1164	1173	1203	1209	1431	1143	1575	612	177	1260	465	390	141
45		Terminal (nt)	1460616	1458196	1462128	1463516	1463934	1465123	1466373	1468548	1471413	1470154	1472907	1474119	1475693	1476294	1476519	1477809	1477929	1478503	1483335
50		Initial (nt)	1458133	5039 1458966	1461157	1462134	1463533	1464083	1465210	1467376	1470211	1471362	1471477	1472977	1474119	1475683	1476343	1476550	1478393	1478892	1483475
		SEO		5039	5040		5042		5044	5045	5046	5047	5048	5049	5050	5051	<del></del>		5054	5055	-
55		SEO	1538	1539			1542	1543	1544	1545	1546	1547	1548	1549	1550	1551	1552	1553	1554	1555	1556

						<del></del>			$\overline{}$	$\overline{}$			_T		[	l	- 1	- 1	- 1		l.	
5		lion		factor IF-2						u	<b>c</b>	P-ammonia	-	ase	ATP-binding	tioning protein or n active irse bacterial	u		ansferase	in	Janit Thase B	*
10		Function	hypothetical protein	translation initiation factor IF-2	hypothetical protein		hypothetical protein	hypothetical protein	DNA repair protein	hypothetical protein	hypothetical protein	CTP synthase (UTP-ammonia ligase)	hypothetical protein	tyrosine recombinase	tyrosin resistance ATP-binding protein	chromosome partitioning protein or ATPase involved in active partitioning of diverse bacterial plasmids	hypothetical protein		thiosulfate sulfurtransferase	hypothetical protein	ribosomal large subunit pseudouridine synthase B	
15		Matched length (a.a.)	84	182	311		260	225	574	394	313	549	157	300	551	258	251		270	(172	229	
20		Similarity (%)	0.99	67.0	60.1		9.69	31.6	63.4	73.1	68.1	76.7	71.3	71.7	59.7	73.6	64.5		67.0	65.7	72.5	
		Identity (%)	61.0	36.3	29.6		38.5	31.6	31.4	41.9	30.4	55.0	36.3	39.7	30.5	44.6	28.3		35.6	33.1	45.9	
25																4						
	Table 1 (continued)	s gene	oniae	ri IF2	дБ		Σ	berculosis	12 recN	berculosis	berculosis	(12 pyrG	ąkG	ureus xerD	diae tIrC	centus par/	bng		a tst	Hnd	En]	
30	Table 1 (c	Homologous gene	Chlamydia pneumoniae	Borrelia burgdorferi IF2	Bacillus subtilis yzgD		Bacillus subtilis yqxC	Mycobacterium tuberculosis H37Rv Rv1695	Escherichia coli K12 recN	Mycobacterium tuberculosis H37Rv Rv1697	Mycobacterium tuberculosis H37Rv Rv1698	Escherichia coli K12 pyrG	Bacillus subtilis yakG	Staphylococcus aureus xerD	Streptomyces fradiae tlrC	Caulobacter crescentus parA	Bacillus subtilis ypuG		Datisca glomerata tst	Bacillus subtilis ypuH	Bacillus subtilis rluB	
<i>35</i>			<u> </u>	ă	+	_	$\vdash$	ΣI	m	≥I	21		+	1	1			ļ	_	1	1	
40		db Match	GSP-Y35814	Sp. IF2 BORBU	sp.YZGD_BACSU		sp.YQXC_BACSU	sp:YFJB_HAEIN	Sp. RECN ECOLI	pir.H70502	pir.A70503	sp:PYRG_ECOLI	Sp.YOKG BACSU	no. AF093548 1	Sp:TLRC_STRFR	gp:CCU87804_4	sp. YPUG_BACSU		qp:AF109156 1	Sp:YPUH BACSU	sp:RLUB_BACSU	
		ORF (bp)	27.3	+-		162	819	873	1779	1191	963	1662	657	9 5	1530	783	765	561	867	543	756	
45		Terminal O	2	_		+-	+	<del></del>	1490881	1	1493109	1495174		1496772	1496795	1499645	1500695	-	<del>-</del>		1504238	
50		Initial	4402006	14835350	1486042	1487032	1487238	1488146	1489103	1490944	1492147	1493513	1405205			5070 1498863	1499931					
		SEQ.	(a.a.)	<del>-</del>					5063		5065				5069	5070	5071	5072	5073	5074	5075	
55				155/		<del></del>		-	1563 5		1565				1569		1571	+	_		<del></del>	

5			-			-						e protein						-2,4-dienoate	SecA subunit	otein		
10			Function	cytidylate kinase	GTP binding protein			methyltransferase	ABC transporter	ABC transporter		hypothetical membrane protein		Na+/H+ antiporter			hypothetical protein	2-hydroxy-6-oxohepta-2,4-dienoate hydrolase	preprotein translocase SecA subunit	signal transduction protein	hypothetical protein	hypothetical protein
15			Matched length (a.a.)	220	435			232	499	602		257		499			130	210	805	132	234	133
20			Similarity (%)	73.6	74.0			67.2	60.1	56.3		73.2		61.5			57.7	63.8	61.7	93.2	74.4	63.2
			Identity (%)	38.6	42.8			36.2	29.7	31.2		39.7		25.7			36.9	25.2	35.2	75.8	41.9	30.8
25	1	utinuea)	gene					rculosis	iatum M82B	iatum M82B		ygiE		C 9372			2 o249#9	dus AF0675	A	egmatis garA	erculosis	erculosis
30	; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;	lable 1 (continued)	Homologous gene	Bacillus subtilis cmk	Bacillus subtilis yphC			Mycobacterium tuberculosis Rv3342	Corynebacterium striatum M82B tetA	Corynebacterium striatum M828 tetB		Escherichia coli K12 ygiE		Bacillus subtilis ATCC 9372 nhaG			Escherichia coli K12 o249#9 ychJ	Archaeoglobus fulgidus AF0675	Bacillus subtilis secA	Mycobacterium smegmatis garA	Mycobacterium tuberculosis H37Rv Rv1828	Mycobacterium tuberculosis H37Rv Rv1828
			db Match	P.KCY BACSU	BACSU			sp:YX42_MYCTU	prf.2513302B	orf:2513302A		Sp. YGIE ECOLI		gp.AB029555_1			sp:YCHJ_ECOLI	pir.C69334	sp. SECA_BACSU	gp.AF173844_2	sp:Y0DF_MYCTU	sp:Y0DE_MYCTU
			ORF (bp)	95 069	+-	999	498	813 sl	1554 p	1767 p	825	789 s	189	1548 g	186	420	375 s	1164 p	2289 s	429 g	756 s	e33 s
45	·		Terminal (nt)	1504945	1506573	1506662	1507405	1507917	1510366	1512132	1510843	1512977	1514693	1512980	1514974	1515815	1515408	1515799	1519458	1520029	1520945	1521589
50			Initial (nt)	1504256	1505017	1507327	1507902	1508729	1508813	1510366	1511667			1514527	1515159	1515396		1516962	1517170			1520957
			SEQ NO NO	5076	5077	-+	_1	5080	5081	5082	5083	5084		_	5087	5088		5090	5091	<del></del> -	<del></del>	5094
			004	9	15	8	62	8	120	82	83	84	85	86	87	88	83	8	6	92	93	94

5	 ~	vo						,	,	-	case	2-binding protein	dehydrogenase		-	ing protein l	ine protein	ıtor	oort system	oort system	oort ATP-binding	-	
10		Function	hypothetical protein					hemolysin	hemolysin		DEAD box RNA helicase	ABC transporter ATP-binding protein	6-phosphogluconate dehydrogenase	thioesterase		nodulation ATP-binding protein	hypothetical membrane protein	transcriptional regulator	phosphonates transport system permease protein	phosphonates transport system permease protein	phosphonates transport ATP-binding protein		
15		Matched length (a.a.)	178					342	65		374	245	492	121		235	232	277	281	268	250		
20		Similarity (%)	84.3					0.69	65.5		69.5	66.1	99.2	67.8		68.1	76.3	63.9	63.4	62.3	72.0		
		Identity (%)	71.4					33.9	31.4		41.2	34.3	99.0	39.7		39.6	43.1	26.7	29.9	27.2	44.8		
25	Table 1 (continued)	us gene	berculosis					dP	Tbı		hilus herA	berculosis	wum	berculosis		3 nodl	berculosis	12 yfhH	12 phnE	12 phnE	12 phnC		
30	Table 1 (	Homologous gene	Mycobacterium tuberculosis H37Rv Rv1828					Bacillus subtilis yhdP	Bacillus subtilis yhdT		Thermus thermophilus herA	Mycobacterium tuberculosis H37Rv Rv1348	Brevibacterium flavum	Mycobacterium tuberculosis H37Rv Rv1847		Rhizobium sp. N33 nodl	Mycobacterium tuberculosis H37Rv Rv1686c	Escherichia coli K12 yfhH	Escherichia coli K12 phnE	Escherichia coli K12 phnE	Escherichia coli K12 phnC		
35													Br	ΣÏ			ΣÏ						
40		db Match	sp:YODE_MYCTU					sp:YHDP_BACSU	SP:YHDT_BACSU		gp:TTHERAGEN_1	sp:YD48_MYCTU	gsp:W27613	pir.G70664		sp:NODI_RHIS3	pir:E70501	Sp:YFHH_ECOLI	sp:PHNE_ECOLI	sp:PHNE_ECOLI	sp. PHNC_ECOLI		
		ORF (bp)	573	510	1449	600	930	1062	1380	219	1344	735	1476	462	675	741	741	873	846	804	804	210	1050
45		Terminal (nt)	1522343	1522432	1523052	1525973	1524568	1525473	1526534	1528186	1527987	1530220	1530341	1532394	1532996	1533781	1534521	1534529	1535382	1536227	1537030	1538968	1537870
50		Initial (nt)	1521771	1522941	1524500	1525374	1525497	1526534	1527913	1527968	1529330	1529486	1531816	1531933	1532322	1533041	1533781	1535401	1536227	1537030	1537833	1538759	1538919
		SEQ NO.	5095	5096	5097	5098	5099	5100	5101	5102	5103	5104	5105	5106	5107	5108	5109	5110	5111	5112	5113	5114	5115
55		SEQ NO. (DNA)	1595	1596	1597	1598	1599	1600	1601	1602	1603	1604	1605	1606	1607	1608	1609	1610	1611	1612	1613	1614	1615

												- 1			$\neg$	$\neg$		$\neg$				
5	uı		idine kinase	kinase	cyl-phospholipid	4-methyl-o- sermease	yltransferase		nslocation pump										nannose	/ltransferase		
10	Function		phosphomethylpyrimidine kinase	hydoxyethylthiazole kinase	cyclopropane-fatty-acyl-phospholipid synthase	sugar transporter or 4-methyl-o- phthalate/phthalate permease	purine phosphoribosyltransferase	hypothetical protein	arsenic oxyanion-translocation pump membrane subunit	-	hypothetical protein	sulfate permease	hypothetical protein	,				hypothetical protein	dolichol phosphate mannose synthase	apolipoprotein N-acyltransferase		secretory lipase
15	Matched length (a.a.)		262	249	451	468	156	206	361		222	469	97					110	217	527		392
20	Similarity (%)		70.2	77.5	55.0	66.9	59.0	68.5	54.6		83.8	83.6	20.0		·			87.3	71.0	55.6		55.6
	Identity (%)		47.3	46.6	28.6	32.5	36.5	39.8	23.3		62.2	51.8	39.0					71.8	39.2	25.1		23.7
25 (panuita	gene		ium thiD	ium LT2	rculosis	a Pc701	62 gpt	yebN	s4 arsB		color A3(2)	39 ORFA	R9 ORFG					erculosis	ses pombe	2 Int		11
30 Table 1 (continued)	Homologous gene		Salmonella typhimurium thiD	Salmonella typhimurium LT2 thiM	Mycobacterium tuberculosis H37Rv ufaA1	Burkholderia cepacia Pc701 mopB	Thermus flavus AT-62 gpt	Escherichia coli K12 yebN	Sinorhizobium sp. As4 arsB		Streptomyces coelicolor A3(2) SCI7.33	Pseudomonas sp. R9 ORFA	Pseudomonas sp. F					Mycobacterium tuberculosis H37Rv Rv2050	Schizosaccharomyces pombe dpm1	Escherichia coli K12 Int		Candida albicans lip1
40	db Match		Sp.THID_SALTY	SP.THIM_SALTY	pir.H70830	prf.2223339B	prf.2120352B	9	758_2		gp:SCI7_33	qp:PSTRTETC1 6	GP:PSTRTETC1_7					pir:A70945	prf:2317468A	Sp:LNT_ECOLI		gp:AF188894_1
	ORF (bp)	702	1584	804	1314	1386	474	669	966	483	693	1455	426	615	207	189	750	396	810	1635	741	1224
45	Terminal (nt)	1538963	1539820	1542119	1546289	1546307	1547967	1549349	1550398	1550951	1552237	1553972	1553297	1554070	1555067	1554891	1555086	1556771	1557014	1557859	1559497	1560437
50	Initial (nt)	1539664	1541403	1542922	1544976	1547692	1548440			1550469		1552518	L	1554684	1554861	1555079	1555835	1556376	1557823	1559493		
	SEQ NO.	5116	5117	5118	5119	5120	5121	5122	5123	5124	5125	5126	5127		5129	5130	5131	5132	5133	5134		
	O O S	616	617	618	619	620	624	5 5	623	624	625	626	627	628	629	630	631	632	633	634	1635	1636

																			$\neg$
5			ısferase					dipeptidase		helicase	ein translocase								
10		Function	precorrin 2 methyltransferase	precorrin-6Y C5, 15- methyltransferase			oxidoreductase	dipeptidase or X-Pro dipeptidase		ATP-dependent RNA helicase	sec-independent protein translocase protein	hypothetical protein	hypothetical protein	hypothetical protein	hypothetical protein		hypothetical protein	hypothetical protein	hypothetical protein
15		Matched length (a.a.)	291	411			244	382		1030	268	85	317	324	467		61	516	159
20		Similarity (%)	56.7	8.09			75.4	61.3		55.7	62.7	69.4	61.2	64.8	77.3		80.3	74.2	50.0
		Identity (%)	31.3	32.4			54.1	36.1		26.5	28.7	44.7	31.9	32.4	53.1		54.1	48.6	42.0
25	Ę	3	Sis	SI			sis	11		<b>3</b> e			sis		Sis		Sis	sis	E2014
30	Counting of the Continued	Homologous gene	Mycobacterium tubercutosis H37Rv cobG	Pseudomonas denitrificans SC510 cobl.			Mycobacterium tuberculosis H37Rv RV3412	Streptococcus mutans LT11 pepQ		Saccharomyces cerevisiae YJL050W dob1	Escherichia coli K12 tatC	Mycobacterium leprae MLCB2533.27	Mycobacterium tuberculosis H37Rv Rv2095c	Mycobacterium leprae MLCB2533.25	Mycobacterium tuberculosis H37Rv Rv2097c		Mycobacterium tuberculosis H37Rv Rv2111c	Mycobacterium tuberculosis H37Rv Rv2112c	Aeropyrum pernix K1 APE2014
35 40		db Match	pir:C70764	Sp.COBL_PSEDE			sp:YY12_MYCTU	gp.AF014460_1		sp:MTR4_YEAST	sp.TATC_ECOLI	sp:YY34_MYCLE	sp:YY35_MYCTU	sp:YY36_MYCLE	sp:YY37_MYCTU		pir.B70512	pir.C70512	PIR:H72504
		ORF (bp)	774	1278	366	246	738	1137	639	2787	1002	315	981	972	1425	249	192	1542	480
45		Terminal (nt)	1562553	1562525	1564237	1564482	1564565	1565302	1567106	1567117	1569932	1571068	1571506	1572492	1573491	1575205	1574945	1575406	1577806
50		Initial (nt)	1561780	1563802	1563872	1564237	1565302	1566438	1566468	1569903	1570933	1571382	1572486	1573463	1574915	1574957	1575136	1576947	1577327
		S S S S		5138	5139		5141	5142	5143	5144	5145	5146	5147	5148	5149	5150	5151	5152	5153
55		SEO	(JNA)	1638	1639	1640	1641	1642	1643	1644	1645	1646	1647	1648	1649	1650	1651	1652	1653

		(a)		Т									-					ł
5	Ę	chaperone-lik	e)	ase		protein	orotein	yase	transferase	nutase	olate //transferase	C Coper The C	educiase	protein				netase
10	Function	AAA family ATPase (chaperone-like function)	protein-beta-aspartate methyltransferase	aspartyl aminopeptidase	hypothetical protein	virulence-associated protein	quinolon resistance protein	aspartate ammonia-lyase	ATP phosphoribosyltransferase	beta-phosphoglucomutase	5-methyltetrahydrofolate homocysteine methyltransferase	7	alkyi nyaroperoxiue reductase subunit F	arsenical-resistance protein	arsenate reductase	arsenate reductase		cysteinyl-tkNA syntnetase
15	Matched length	545	281	436	269	69	385	526	281	195	1254		366	388	129	123		387
20	Similarity (%)	78.5	79.0	67.2	71.4	72.5	61.0	99.8	97.5	63.1	62.4		49.5	63.9	64.3	75.6		64.3
	Identity (%)	51.6	57.3	38.1	45.4	40.6	21.8	8 66	8.96	30.8	31.6		22.4	33.0	32.6	47.2		35.9
25	3	arc	닏		Sis	198	orA23	cum MJ233	uno	888	I		s ahpF	ae	olasmid	osis		S
30 F	Homologous gene	Rhodococcus erythropolis arc	Mycobacterium leprae pimT	Homo sapiens	Mycobacterium tuberculosis H37Rv Rv2119	Dichelobacter nodosus A198 vapl	Staphylococcus aureus norA23	Corynebacterium glutamicum (Brevibacterium flavum) MJ233 aspA	Corynebacterium glutamicum ASO19 hisG	Thermotoga maritima MSB8 TM1254	Escherichia coli K12 metH		Xanthomonas campestris ahpF	Saccharomyces cerevisiae S288C YPR201W acr3	Staphylococcus aureus plasmid pl258 arsC	Mycobacterium tuberculosis H37Rv arsC		Escherichia coli K12 cysS
40	db Match	nrf 24223820		50 1	pir.B70513	sp.VAPI_BACNO	prf:2513299A	sp:ASPA_CORGL	gp:AF050166_1	pir:H72277	sp:METH_ECOLI		sp.AHPF_XANCH	sp:ACR3_YEAST	sp.ARSC_STAAU	pir.G70964		sp:SYC_ECOLI
	ORF	(ob)	834	1323	834	264	1209	1578	843	693	3663	570	1026	1176	420	639	378	1212
<b>45</b>	Terminal	(nt)	1578567	1579449	1581640	1582114	1582273	1583913	1585603	1586812	1587573	1591912	1591941	1594512	1594951	1595668	1595844	1
50	Initial	(nt)	1579400	1580771	1580807	1581851	1583481	1585490	1586445	1587504	1591235	1591343	1592966	1593337	1594532	1595030	1596221	1597460
	SEQ	(a.a.)	5155			5158	5150	<del></del>	5161	5162	5163	5164	_	5166	5167	5168	5169	5170
55	SEQ		1655	1656	1657	1658	1850	1660	1661	1662	1663	1664	1665	1666	1667	1668	1669	1670

5	L.	protein			frogenase	-				otin biosynthetic	des predicted to or vaccines and		-				ransferase	, and ne and ystem kinase	mutase alpha
10	Function	bacitracin resistance protein	oxidoreductase	lipoprotein	dihydroorotate dehydrogenase			transposase		bio operon ORF I (biotin biosynthetic enzyme)	Neisserial polypeptides predicted to be useful antigens for vaccines and diagnostics		ABC transporter		ABC transporter		puromycin N-acetyltransferase	LAO(lysine, arginine, and ornithine)/AO (arginine and ornithine)transport system kinase	methylmalonyl-CoA mutase alpha subunit
15	Matched length (a.a.)	255	326	359	334			360		152	198		265		535		99	339	741
20	Similarity (%)	69.4	62.6	53.5	67.1			55.3		75.0	33.0		68.7		67.1		56.4	72.3	87.5
	Identity (%)	37.3	33.4	27.0	44.0			34.7		44.1	26.0		43.6		36.8		32.4	43.1	72.2
25 (pən	60	A	sus	osis	3			tnpA		ā			m M82B		m M82B		pac	え	nensis
© Table 1 (continued)	Homologous gene	Escherichia coli K12 bacA	Agrobacterium tumefaciens mocA	Mycobacterium tuberculosis H37Rv lppL	Agrocybe aegerita ura1			Pseudomonas syringae tnpA		Escherichia coli K12 ybhB	Neisseria meningitidis		Corynebacterium striatum M82B tetB		Corynebacterium striatum M82B tetA	•	Streptomyces anulatus pac	Escherichia coli K12 argK	Streptomyces cinnamonensis A3823.5 mutB
40	db Match	sp:BACA_ECOLI E	prf.2214302F	pir:F70577	SP. PYRD_AGRAE			gp:PSESTBCBAD_1		sp:YBHB_ECOLI	GSP:Y74829		prf.2513302A		prf.2513302B		pir.JU0052	ECOLI	sp:MUTB_STRCM
	ORF (bp)	879	948	666	1113	351	807	1110	486	531	729	603	1797	249	1587	351	609	1089	2211
45	Terminal (nt)	1597745	1599614	1600677	1601804	1601931	1603466	1604629	1604830	1605281	1606689	1608248	1605861	1609335	1607661	1609842	1610844	1611150	1612234
50	Initial (nt)	1598623	1598667	1599679	1600692	1602281	1602660	1603520	1605315	1605811	1605961	1607646	1607657	1609087	1609247	1610192	1610236		1614444
	SEQ NO.	+	5172	5173	5174	5175	5176	5177	5178	5179	5180	5181	5182	5183	5184	5185			5188
55	SEQ.	<del></del>	1672	1673	1674	1675	1676	1677	1678	1679	1680	1681	1682	1683	1684	1685	1686	1687	1688

5	E	lutase beta	ne protein		ne protein	ne protein							ıtor					
10	Function	methylmalonyl-CoA mutase beta subunit	hypothetical membrane protein		hypothetical membrane protein	hypothetical membrane protein	hypothetical protein		ferrochelatase	invasin		aconitate hydratase	transcriptional regulator	GMP synthetase	hypothetical protein	hypothetical protein		hypothetical protein
15	Matched length		224		370	141	261		364	611		959	174	235	221	98		446
20	Similarity (%)	68.2	70.1		87.0	78.7	72.8		65.7	56.5		85.9	81.6	51.9	62.0	80.2		96.1
	Identity (%)	41.6	39.7		64.1	44.7	51.0		36.8	25.5		6.69	54.6	21.3	32.6	37.2		61.2
<i>25</i>		ensis	sis		sis	osis	A3(2)		enreichii			osis	losis	景	A3(2)	chii		4C58
20 September 1 (continued)	Homologous gene	innamone	tuberculc		tubercul	tubercul	oelicolor		um freud inii hemH	faecium		tubercu	tubercu tc	s jannas	coelicolo	ıs jannas		ingitidis
30 E	SolomoH	Streptomyces cinnamonensis A3823.5 mutA	Mycobacterium tuberculosis H37Rv Rv1491c		Mycobacterium tuberculosis H37Rv Rv1488	Mycobacterium tuberculosis H37Rv Rv1487	Streptomyces coelicolor A3(2) SCC77.24		Propionibacterium freudenreichii subsp. Shermanii hemH	Streptococcus faecium		Mycobacterium tuberculosis H37Rv acn	Mycobacterium tuberculosis H37Rv Rv1474c	Methanococcus jannaschii MJ1575 guaA	Streptomyces coelicolor A3(2) SCD82.04c	Methanococcus jannaschii MJ1558		Neisseria meningitidis MC58 NMB1652
35	-		1			ΣI	SS	-	<del>                                     </del>			Z T	2 -		0, 0,			Ф.
40	db Match	sp:MUTA_STRCM	sp:YS13_MYCTU		sp:YS09_MYCTU	pir:B70711	gp:SCC77_24		sp:HEMZ_PROFR	sp:P54_ENTFC		pir.F70873	pir.E70873	pir:F64496	gp:SCD82_4	pir.E64494		gp:AE002515_
	ORF		723	597	1296	435	843	783	1110	1800	498	2829	564	756	663	267	393	1392
45	<u>a</u>	1614451	1617300	1617994	1618321	1619672	1620167	1621838	1621841	1623027	1625428	1629107	1629861	1630668	1630667	1631926	1631353	1633324
50	Initial	(nt) 1616298	1616578	1617398	1619616	1620106	1621009	1621056		1624826			1629298	1629913	1631329	1631660	1631745	
	SEQ			5191		5193	5194	5105	5196	5197	5198	5199	5200	5201	5202	5203		
55	SEQ		<del>i</del>	1691	<del></del>	1693	1694	1605		1697	169R	1699	1700	1701	1702	1703	1704	1705

-	. [		Т	T							$\prod$	tein		tein							ase		
5		_			Pase P							sed lipopro		binding pro				ragment			nose reduct	in	
10		Function	antigenic protein	antigenic protein	cation-transporting ATPase P		hypothetical protein					host cell surface-exposed lipoprotein	integrase	ABC transporter ATP-binding protein		sialídase	transposase (IS1628)	transposase protein fragment	hypothetical protein		dTDP-4-keto-L-rhamnose reductase	nitrogen fixation protein	
15		Matched length (a.a.)	113	152	883		120					107	154	497		387	236	37	88		107	149	
20		Similarity (%)	0.09	0.69	73.2		58.3					73.8	60.4	64.4		72.4	100.0	72.0	43.0		70.1	85.2	
		Identity (%)	54.0	59.0	42.6		35.8					43.0	34.4	32.8		51.9	9.66	64.0	32.0		32.7	63.8	1
25	ntinued)	gene	se ORF24	3e	CC6803		olor A3(2)					sophilus	nt	yjjK		idifaciens	utamicum \G1 tnpB	utamicum			Orsay	ae	
30	Table 1 (continued)	Homologous gene	Neisseria gonorrhoeae	Neisseria gonorrhoeae	Synechocystis sp. PCC6803 sll1614 pma1		Streptomyces coelicolor A3(2) SC3D11.02c					Streptococcus thermophilus phage TP-J34	Corynephage 304L int	Escherichia coli K12 yjjK		Micromonospora viridifaciens ATCC 31146 nedA	Corynebacterium glutamicum 22243 R-plasmid pAG1 tnpB	Corynebacterium glutamicum TnpNC	Plasmid NTP16		Pyrococcus abyssi Orsay PAB1087	Mycobacterium leprae MLCL536.24c nifU7	
40		db Match	GSP: Y38838		Z Z Z		gp:SC3D11_2					prf:2408488H	prf.2510491A	sp:YJJK_ECOLI		sp:NANH_MICVI	gp:AF121000_8	GPU:AF164956_23	GP:NT1TNIS_5		pir.B75015	pir.S72754	
		ORF (bp)	480	456	2676	783	489	1362	357	156	162	375	456	1629	1476	1182	708	243	261	585	423	447	
45		Terminal (nt)	1632109	1632682	<del>                                     </del>	1633781	1636244	1638442	1638776	1639520	1639817	1640155	1641001	1641046	1642743	1644318	1646368	1646063	1645601	1647133	1647212	1647651	
50		Initial (nt)	g	-		1634563	1636732	1637081	1639132		1639656		1640546			1645499	1645661	1645821	1645861	1646549		1648097	
		SEO	(a.a.)			5209		5211			5214		5216	5217	5218	5219	5220	5221	5222	5223	5224	5225	
			<b>E S</b>			60	+	71	712	713	714	715	716	717	718	719	720	721	722	723	1724	1725	

Table 1 (continued)   Continued   Contin			50	45		40	· <b>35</b>	30	<b>25</b> .	20	15	5
SEQ         Initial         Terminal         ORF         db Match         Homologous gene         Identity (%)         Match (%)         (%)							Tat	ile 1 (continued)				
5228         1648548         1648709         162         PIR:C72506         Aeropyuum pernix K1 APE2025         48.0         57.0         52.0           5227         1448361         1648100         1263         pir.S72781         Mycobacterium leprae nifs         64.7         84.4         411           5228         165012         1649367         756         gp.SCC22_4         Streptomyces coelicolor A3(2)         70.2         89.3         252           5228         165012         1649367         756         gp.SCC22_4         Streptomyces coelicolor A3(2)         70.2         89.3         377           5229         1651424         1650249         1176         pir.A70872         Hycobacterium tuberculosis         55.2         83.0         377           5231         1652896         1651403         1443         sp.Y074_SYNY3         Streptomyces coelicolor A3(2)         46.1         71.4         217           5231         1652896         16514043         1652894         693         pir.S72783         Mycobacterium tuberculosis         50.2         77.3         317           5234         1655671         1658675         804         pir.S72783         Mycobacterium teprae         41.0         74.6         291           <	<del></del>	SEQ.	Initial (nt)	Terminal (nt)	ORF (bp)	db Match	Ног	eueg snogolou	Identity (%)		<b></b>	Function
5227         1649362         1648100         1263         pir.S72761         Mycobacterium leprae nils         64.7         84.4         411           5228         1650122         1649367         756         gp.SCC22_4         Streptomyces coelicolor A3(2)         70.2         89.3         252           5229         1651424         1650249         1176         pir.A70872         Hycobacterium luberculosis         55.2         83.0         377           5229         1651424         1650249         1176         pir.A70872         Hycobacterium luberculosis         55.2         83.0         377           5231         1652875         165143         1443         sp.YO74_SYNY3         Streptomyces coelicolor A3(2)         46.1         71.4         217           5231         1652876         1652894         693         gp.SCC22_8         Streptomyces coelicolor A3(2)         46.1         71.4         217           5232         1654043         165287         1629         pir.F70871         Hycobacterium luberculosis         36.3         67.8         518           5234         165677         1657515         804         pir.S72778         Mycobacterium luberculosis         41.0         74.8         291           5237	728	5226		1648709	162	PIR:C72506	Aeropyrum	pernix K1 APE20		57.0	52	hypothetical protein
5228         1650122         1649367         756         gp:SCC22_4         Streptomyces coelicolor A3(2)         70.2         89.3         252           5228         1650122         1650124         1176         pir.A70872         HXPobacterium luberculosis         55.2         83.0         377           5220         1651424         1650249         1176         pir.A70872         HXPobacterium luberculosis         55.2         83.0         377           5230         1652894         693         gp:SCC22_8         Streptomyces coelicolor A3(2)         46.1         71.4         217           5231         1652894         693         gp:SCC22_8         Streptomyces coelicolor A3(2)         46.1         71.4         217           5231         1653686         1652894         693         gp:SCC22_8         Streptomyces coelicolor A3(2)         46.1         71.4         217           5231         1653681         165671         1620         pir.F70871         HXPODAGCERIum luberculosis         36.2         77.3         317           5234         1656712         1657515         804         pir.C70871         Mycobacterium leprae         50.2         77.8         50           5236         1655140         357         804 <td></td> <td>5227</td> <td>_</td> <td>1648100</td> <td>1263</td> <td><del></del></td> <td>Mycobacter</td> <td>rium leprae nifS</td> <td></td> <td>84.4</td> <td>411</td> <td>nitrogen fixation protein</td>		5227	_	1648100	1263	<del></del>	Mycobacter	rium leprae nifS		84.4	411	nitrogen fixation protein
5229         1651424         1650249         1176         pir.A70872         Mycobacterium tuberculosis         55.2         83.0         377         1           5230         1652875         1651433         1443         sp:Y074_SYNY3         Synechocystis sp. PCC6803         41.0         73.0         493         .           5231         1652876         1652894         693         gp:SCC22_8         Streethocystis sp. PCC6803         46.1         71.4         217         1           5232         165681         1656701         1629         pir.F70871         Mycobacterium tuberculosis         36.3         67.8         518         1           5234         165681         165670         1020         pir.S72783         Mycobacterium leprae         50.2         77.3         317         .           5235         165681         165707         1658675         804         pir.C70871         Mycobacterium leprae         41.0         74.8         266           5235         165681         1651040         357         Mycobacterium leprae         41.0         74.8         267           5236         165681         1651404         357         804         Pir.C70871         Mycobacterium leprae         41.0         74.8<		5228		1649367	756	<del></del>	Streptomyo SCC22.04c	es coelicolor A3(2		89.3	252	ABC transporter ATP-binding protein
5230         1652875         1651433         1443         sp.Y074_SYNY3         Synechocystis sp. PCC6803         41.0         73.0         493           5231         1652875         1652894         693         gp:SCC22_8         Streptomyces coelicolor A3(2)         46.1         71.4         217         1           5232         1654043         1655671         1629         pir.F70871         Mycobacterium tuberculosis         36.3         67.8         518           5233         1655681         165670         1020         pir.S72788         Mycobacterium leprae         41.0         74.8         286           5234         165671         1658675         804         pir.S72778         Mycobacterium reprae         41.0         74.6         291           5236         1659496         1659140         357         Mycobacterium reprae         43.0         74.6         291           5237         1669508         1661138         1629         pir.C71156         Pyrococcus horikoshii PH0450         23.4         51.0         418           5238         1663598         1662630         969         pir.C71156         Pyrococcus horikoshii PH0450         37.5         70.9         323           5240         1664403	1	5229	_!	1650249	1176		Mycobacte H37Rv Rv1	rium tuberculosis 1462	55.2		377	hypothetical protein
5231         1653686         1652894         693         gp:SCC22_8         Streptomyces coelicolor A3(2)         46.1         71.4         217           5232         1654043         1655671         1629         pir.F70871         Mycobacterium tuberculosis         36.3         67.8         518           5232         165581         1656700         1020         pir.S72783         Mycobacterium leprae         50.2         77.3         317           5234         1656712         1657515         804         pir.S7278         Mycobacterium leprae         41.0         74.8         266           5236         1657677         1658675         999         pir.C70871         Mycobacterium leprae         43.0         74.6         291           5236         1657677         1658675         999         pir.C70871         Mycobacterium tuberculosis         43.0         74.6         291           5236         1657677         1658676         999         pir.C7186         Pyrococcus horikoshii PH0450         23.4         51.0         418           5238         1661578         1662532         975         sp. QOR_ECOL1         Escherichia coli K12 qor         37.5         70.9         323           5240         1666502         <		5230		1651433	1443		Synechocy slr0074	stis sp. PCC6803	41.0		493	ABC transporter
5232         1654043         1656671         1629         pir.F70871         Mycobacterium tuberculosis         36.3         67.8         518           5233         1655681         1656700         1020         pir.S72783         Mycobacterium leprae         50.2         77.3         317           5234         1656712         1657515         804         pir.S72778         Mycobacterium leprae         41.0         74.8         266           5235         1657677         1658675         999         pir.C70871         Mycobacterium tuberculosis         43.0         74.6         291           5236         1653496         1659140         357         Mycobacterium tuberculosis         43.0         74.6         291           5237         1653658         1661136         1629         pir.C71456         Pyrococcus horikoshii PH0450         23.4         51.0         418           5238         1661578         1662630         969         gp.MVCOXABC_3         Nitrobacter winogradskyi coxC         37.6         66.8         295           5240         1664403         1666502         2100         gp.AB023377_1         ATCC 31833 tkt         100.0         100.0         67.5           5241         1666673         1667752		5231		1652894	693	gp:SCC22_8	Streptomyc SCC22.08c	ses coelicolor A3(2			217	DNA-binding protein
5233         1655681         1656700         1020         pir.S72783         Mycobacterium leprae         50.2         77.3         317           5234         1656712         1657515         804         pir.S72778         Mycobacterium leprae         41.0         74.8         266           5235         1657677         1658675         999         pir.C70871         Mycobacterium leprae         43.0         74.6         291           5236         1659496         1659140         357         Mycobacterium tuberculosis         43.0         74.6         291           5237         1659608         1661136         1629         pir.C71156         Pyrococcus horikoshii PH0450         23.4         51.0         418           5238         1661578         1662552         975         sp. QOR_ECOLI         Escherichia coli K12 qor         37.5         70.9         323           5239         1661578         1662552         975         sp. QOR_ECOLI         Escherichia coli K12 qor         37.5         66.8         295           5240         1664403         1662630         969         gp. NWCOXABC_3         Nitrobacterium glutamicum         100.0         100.0         675           5241         1666673         168752		5232		1655671	1629		Mycobacte H37Rv Rv	rium tuberculosis 1459c	36.3		518	hypothetical membrane protein
5234         1656712         1657515         804         pir.S72778         Mycobacterium leprae         41.0         74.8         266           5235         1657677         1658675         999         pir.C70871         Mycobacterium tuberculosis         43.0         74.6         291           5236         1659496         1659140         357         Mycobacterium tuberculosis         43.0         74.6         291           5237         1659508         1661136         1629         pir.C71156         Pyrococcus horikoshii PH0450         23.4         51.0         418           5238         1661578         166136         1629         975         sp.QOR_ECOLI         Escherichia coli K12 qor         37.5         70.9         323           5239         1663598         1662630         969         gp.NWCOXABC_3         Nitrobacter winogradskyi coxC         37.6         66.8         295           5240         1664403         1666502         2100         gp.AB023371         ATCC 31833 tkt         100.0         675           5241         1666673         1667764         1666601         1164         Mycobacterium leprae         62.0         85.2         358		5233		1656700	1020		Mycobacte MLCL536.	rium leprae 31 abc2	50.2		317	ABC transporter
5235         1658675         999         pir.C70871         Mycobacterium tuberculosis         43.0         74.6         291           5236         1659496         1659140         357         Priconaccus horikoshii PH0450         23.4         51.0         418           5237         1659508         1661378         1629         pir.C71156         Pyrococcus horikoshii PH0450         23.4         51.0         418           5237         1659508         1661378         162552         975         sp:QOR_ECOLI         Escherichia coli K12 qor         37.5         70.9         323           5239         1663598         1662530         969         gp:NWCOXABC_3         Nitrobacter winogradskyi coxC         37.6         66.8         295           5240         1664403         1666502         2100         gp:AB023377_1         ATCC 31833 tkt         100.0         100.0         675           5241         1666673         1667752         1080         sp:TAL_MYCLE         Mycobacterium leprae         62.0         85.2         358	1	5234		1657515	804	pir:S72778	Mycobacte MLCL536.	rium leprae 32	41.0		266	hypothetical protein
5236         1659496         1659140         357         Pyrococcus horikoshii PH0450         23.4         51.0         418           5237         1659508         1661136         1629         pir.C71156         Pyrococcus horikoshii PH0450         23.4         51.0         418           5238         1661578         1662552         975         sp.QOR_ECOLI         Escherichia coli K12 qor         37.5         70.9         323           5239         1663598         1662630         969         gp.NWCOXABC_3         Nitrobacter winogradskyi coxC         37.6         66.8         295           5240         1664403         1666502         2100         gp:AB023377_1         ATCC 31833 tkt         100.0         100.0         675           5241         1666673         1667752         1080         sp:TAL_MYCLE         Mycobacterium leprae         62.0         85.2         358           5242         1667764         1666601         1164         ATC         ATCL536.39 tal         39 tal		5235		1658675	666	pir.C70871	Mycobacte H37Rv Rv	rium tuberculosis 1456c	43.0		291	hypothetical protein
5237         1659508         1661136         1629         pir.C71156         Pyrococcus horikoshii PH0450         23.4         51.0         418           5238         1661578         1662552         975         sp:QOR_ECOLI         Escherichia coli K12 qor         37.5         70.9         323           5239         1663598         1662630         969         gp:NWCOXABC_3         Nitrobacter winogradskyi coxC         37.6         66.8         295           5240         1664403         1666502         2100         gp:AB023377_1         ATCC 31833 tkt         100.0         100.0         675           5241         1666673         1667752         1080         sp:TAL_MYCLE         MLCL536.39 tal         62.0         85.2         358           5242         1667764         1666601         1164         164         164         166         165         166         <		_		1659140	357							
5238         1661578         1662552         975         sp:QOR_ECOLI         Escherichia coli K12 qor         37.5         70.9         323           5239         1663598         1662630         969         gp:NWCOXABC_3         Nitrobacter winogradskyi coxC         37.6         66.8         295           5240         1664403         1666502         2100         gp:AB023377_1         ATCC 31833 tkt         100.0         100.0         675           5241         1666673         1667752         1080         sp:TAL_MYCLE         Mycobacterium leprae         62.0         85.2         358           5242         1667764         1666601         1164         n         n         n         n         n		+		╄	1629	+	Pyrococcu	s horikoshii PH04	-		418	helicase
5239         1663598         1662630         969         gp:NWCOXABC_3         Nitrobacter winogradskyi coxC         37.6         66.8         295           5240         1664403         1666502         2100         gp:AB02337_1         Corynebacterium glutamicum         100.0         100.0         675           5241         1666673         1667752         1080         sp:TAL_MYCLE         Mycobacterium leprae         62.0         85.2         358           5242         1667764         1666601         1164         1164         1164         1164		<del></del>			975	<del>-</del> -	Escherichi	a coli K12 qor	37.5	$\dashv$	323	quinone oxidoreductase
5240         1664403         1666502         2100         gp:AB02337_1         Corynebacterium glutamicum         100.0         100.0         675           5241         1666673         1667752         1080         sp:TAL_MYCLE         Mycobacterium leprae         62.0         85.2         358           5242         1667764         1666601         1164 <td< td=""><td></td><td></td><td></td><td>1</td><td>696</td><td><del> </del></td><td></td><td>r winogradskyi co&gt;</td><td></td><td>-</td><td>295</td><td>cytochrome o ubiquinol oxidase assembly factor / heme O synthase</td></td<>				1	696	<del> </del>		r winogradskyi co>		-	295	cytochrome o ubiquinol oxidase assembly factor / heme O synthase
5241         1666673         1667752         1080         sp:TAL_MYCLE         Mycobacterium leprae         62.0         85.2         358           5242         1667764         1666601         1164         1164         1164         1164         1164         1164         1164         1164         1164         1164         1164         1164         1164         1164         1164         1164         1164         1164         11666				1	2100	<del>  -</del>	Corynebac ATCC 318	sterium glutamicur 133 tkt	-			transketolase
5242 1667764 1666601 116		<del></del>			1080		Mycobacte MLCL536.	erium leprae 39 tal	62.0		358	transaldolase
			1667764		116							

5				se 6-	ıase			-				rase	otein	es	phate				unit C
10		Function	glucose-6-phosphate dehydrogenase	oxppcycle protein (glucose 6- phosphate dehydrogenase assembly protein)	6-phosphogluconolactonase	sarcosine oxidase	transposase (IS1676)	sarcosine oxidase				triose-phosphate isomerase	probable membrane protein	phosphoglycerate kinase	glyceraldehyde-3-phosphate dehydrogenase	hypothetical protein	hypothetical protein	hypothetical protein	excinuclease ABC subunit C
15		Matched length (a.a.)	484	318	258	128	200	205				259	128	405	333	324	309	281	701
20		Similarity (%)	100.0	71.7	58.1	57.8	46.6	100.0				9.66	51.0	98.5	99.7	87.4	82.5	76.2	61.5
		Identity (%)	93.8	40.6	28.7	35.2	24.6	100.0				99.2	37.0	98.0	99.1	63.9	56.3	52.0	34.4
25	(panu	- ue		losis	siae		olis	micum				micum iA	isiae	ımicum gk	amicum ap	ulosis	ulosis	ulosis	C6803
<i>30</i>	Table 1 (continued)	Homologous gene	Brevibacterium flavum	Mycobacterium tuberculosis H37Rv Rv1446c opcA	Saccharomyces cerevisiae S288C YHR163W sol3	Bacillus sp. NS-129	Rhodococcus erythropolis	Corynebacterium glutamicum ATCC 13032 soxA				Corynebacterium glutamicum AS019 ATCC 13059 tpiA	Saccharomyces cerevisiae YCR013c	Corynebacterium glutamicum AS019 ATCC 13059 pgk	Corynebacterium glutamicum AS019 ATCC 13059 gap	Mycobacterium tuberculosis H37Rv Rv1423	Mycobacterium tuberculosis H37Rv Rv1422	Mycobacterium tuberculosis H37Rv Rv1421	Synechocystis sp. PCC6803 uvrC
40		db Match	gsp:W27612	pir.A70917	sp. SOL3_YEAST	sp:SAOX_BACSN	gp.AF126281_1	gp:CGL007732_5				sp:TPIS_CORGL	SP:YCQ3_YEAST	sp:PGK_CORGL	sp:G3P_CORGL	pir.D70903	sp:YR40_MYCTU	sp:YR39_MYCTU	sp:UVRC_PSEFL
		ORF (bp)	1452	957	705	405	1401	840	174	687	981	777	408	1215	1002	981	1023	927	2088
45		Terminal (nt)	1669401	1670375	1671099	1671273	1673123	1673266	1677384	1678070	1680128	1680332	1681670	1681190	1682624	1684117	1685110	1686152	1687103
50		Initial (nt)	1667950	1669419	1670395	1671677			1677211		1679148		1681263	1682404	1683625	1685097	1686132	1687078	1689190
		SEO	(a.a.) 5243	5244	5245	5246	5247	5248	5249	5250	5251	5252	5253	5254	5255	5256	5257	5258	5259
		g o	43 (A)	44	45	46	747	48	749	135	751	752	753	754	755	756	757	758	759

·r			T		1	Т			$\top$					T	1			T	
5	c	-	ımazıne	by rib operon	protein	by rib operon	ll and 3, 4- e 4-phosphate ynthesis)	pha chain	aminase	-epimerase	L1/NOP2	yltransferase	ase	·	e synthetase	netabolism			
10	Function	hypothetical protein	6,7-dimethyl-8-ribityllumazine synthase	polypeptide encoded by rib operon	riboflavin biosynthetic protein	polypeptide encoded by rib operon	GTP cyclohydrolase II and 3, 4- dihydroxy-2-butanone 4-phosphate synthase (riboflavin synthesis)	riboflavin synthase alpha chain	riboflavin-specific deaminase	ribulose-phosphate 3-epimerase	nucleolar protein NOL1/NOP2 (eukaryotes) family	methionyl-tRNA formyltransferase	polypeptide deformylase	primosomal protein n	S-adenosylmethionine synthetase	DNA/pantothenate metabolism flavoprotein	hypothetical protein	guanylate kinase	integration host factor
15	Matched length (a.a.)	150 h	154 . 8		Ţ	106 P	404	211	365	234 г	448	308	150		407	409	18	186	103
20	Similarity N	68.7	72.1	68.0	48.0	52.0	84.7	79.2	62.7	73.1	2.09	67.9	72.7	46.3	99.5	6.08	87.7	74.7	90.3
	Identity S	32.7	43.5	29.0	26.0	44.0	65.6	47.4	37.3	43.6	30.8	41.6	44.7	22.9	99.3	58.0	70.4	39.8	80.6
25 (Pa		sis					sis ribA	78 ribE		ae		sa fmt			AJ-233	osis	osis	iae guk1	losis
os Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv1417	Escherichia coli K12	Bacillus subtilis	Bacillus subtilis	Bacillus subtilis	Mycobacterium tuberculosis ribA	Actinobacillus pleuropneumoniae ISU-178 ribÉ	Escherichia coli K12 ribD	Saccharomyces cerevisiae S288C YJL121C rpe1	Escherichia coli K12 sun	Pseudomonas aeruginosa fmt	Bacillus subtilis 168 def	Escherichia coli priA	Brevibacterium flavum MJ-233	Mycobacterium tuberculosis H37Rv RV1391 dfp	Mycobacterium tuberculosis H37Rv Rv1390	Saccharomyces cerevisiae guk1	Mycobacterium tuberculosis H37Rv Rv1388 mIHF
35		Mycoba H37Rv	Escheri	Bacillus	Bacillus	Bacillus	Мусоря	Actinot	Escher	Sacchi S288C	Escher	Pseudi	Bacillu	Eschei	Brevib	Mycob H37Rv	Mycob H37Rv	Sacch	Mycob H37R\
40	db Match	sp:YR35_MYCTU	sp.RISB_ECOLI	GSP:Y83273	GSP:Y83272	GSP:Y83273	gp.AF001929_1	SP.RISA_ACTPL	sp.RIBD_ECOLI	sp:RPE_YEAST	sp.SUN_ECOLI	SD:FMT PSEAE	SP:DEF BACSU	Sp:PRIA_ECOLI	gsp:R80060	sp:DFP_MYCTU	sp:YD90_MYCTU	pir.KIBYGU	pir.B70899
	ORF (bp)	579	477	228	714	336	1266	633	984	657	1332	945	507	2064	1221	1260	291	627	318
45	Terminal (nt)	1689201	1689869	1690921	1691421	1691347	1690360	1691639	1692275	1693262	1693967	1695499	1696466	1697084	1699177	1700508	1702032	1702411	1702991
50	Initial (nt)	1689779	1690345	1690694	-	<del></del>	1691625	1692271	1693258		1695298	1696443				1	1702322	1703037	1703308
	SEQ	<del>-i</del>	5261	5262				5266	5267	5268	5269	5270	5271	5272	5273	5274	5275	5276	5277
55		1760	1761	1762	<del></del>		<del></del>	1766	1767		_		1771	1772	1773	1774	1775	1776	1777

	_																	
5		Function	orotidine-5'-phosphate decarboxylase	carbamoyl-phosphate synthase large chain	carbamoyl-phosphate synthase small chain	dinydroorotase	aspartate carbamoyltransferase	phosphoribosyl transferase or pyrimidine operon regulatory protein	cell division inhibitor				N utilization substance protein B (regulation of rRNA biosynthesis by transcriptional antitermination)	elongation factor P	cytoplasmic peptidase	3-dehydroquinate synthase	shikimate kinase	type IV prepilin-like protein specific leader peptidase
15		Matched length (a.a.)	276	1122	381	402	311	176	297				137	187	217	361	166	142
20		Similarity (%)	73.6	77.5	70.1	67.7	79.7	1.08	73.4				69.3	98.4	100.0	99.7	100.0	54.9
		Identity (%)	51.8	53.1	45.4	42.8	48.6	54.0	39.7				33.6	97.9	99.5	98.6	100.0	35.2
25 1	lable I (confinded)	ıs gene	berculosis	ara	uginosa	us DSM 405	uginosa	us DSM 405	berculosis				ısB	ctofermentum	glutamicum	glutamicum	glutamicum	phila tapD
·	lable 1 (i	Homologous gene	Mycobacterium tuberculosis H37Rv uraA	Escherichia coli carB	Pseudomonas aeruginosa ATCC 15692 carA	Bacillus caldolyticus DSM 405 pyrC	Pseudomonas aeruginosa ATCC 15692	Bacillus caldolyticus DSM 405 pyrR	Mycobacterium tuberculosis H37Rv Rv2216				Bacillus subtilis nusB	Brevibacterium lactofermentum ATCC 13869 efp	Corynebacterium glutamicum AS019 pepQ	Corynebacterium glutamicum AS019 aroB	Corynebacterium glutamicum AS019 aroK	Aeromonas hydrophila tapD
40		db Match	SP.DCOP_MYCTU	pir.SYECCP E	SP.CARA_PSEAE	Sp.PYRC_BACCL p	Sp.PYRB_PSEAE	Sp.PYRR_BACCL B	Sp.Y00R_MYCTU				sp.NUSB_BACSU E	sp.EFP_BRELA	gp:AF124600_4	gp:AF124600_3	gp:AF124600_2	sp.LEP3_AERHY
		ORF (bp)	834 sp	3339 pir	1179 sp	1341 sp	936 sp	576 sp	1164 sp	477	462	210	681 sp	561 sp	1089 gp	1095 gp	492 gp	411 sp
<b>45</b>		Terminal (nt)	1703517	1704359	1707706	1709017	1710413	1711352	1713759	1714306	1714760	1714950	1715382	1716132	1716780	1717938	1719107	1720971
50		Initial (nt)	1704350	1707697	1708884	1710357	1711348	1711927	1712596	1713830	1714299	1714741	1716062	1716692	1717868	1719032	1719598	1721381
		SEQ NO.		5279	5280	5281	5282	5283	5284	5285	5286	5287	5288	5289	5290	5291	5292	5293
		O O S	78	779	80	781.	782	783	84	785	98	787	788	789	790	791	792	793

5

ſ							$\neg \tau$					$\top$	$\top$	$\neg$		T	$\neg$	
	Function	bacterial regulatory protein, arsR family	ABC transporter		iron(III) ABC transponer, periplasmic-binding protein	ferrichrome transport ATP-binding protein	shikimate 5-dehydrogenase	hypothetical protein	hypothetical protein	alanyi-tRNA synthetase	hypothetical protein		aspartyl-tRNA synthetase	hypothetical protein	glucan 1,4-alpha-glucosidase	phage infection protein		transcriptional regulator
	Matched length (a.a.)	83	340		373	230	259	395	161	894	454		591	297	839	742		192
	Similarity (%)	68.7	73.2		50.7	71.7	60.0	70.1	9.69	71.8	84.8		89.2	74.1	53.6	54.0		62.0
	Identity (%)	45.8	35.9		23.6	38.3	50.0	41.8	52.8	43.3	65.4		7	46.1	26.1	23.1		29.2
Table 1 (confinued)	Homologous gene	Streptomyces coelicolor A3(2) SC1A2.22	Corynebacterium diphtheriae hmuU		Pyrococcus abyssi Orsay PAB0349	Bacillus subtilis 168 fhuC	Mycobacterium tuberculosis H37Rv aroE	Mycobacterium tuberculosis H37Rv Rv2553c	Mycobacterium tuberculosis H37Rv Rv2554c	Thiobacillus ferrooxidans ATCC 33020 alaS	Mycobacterium tuberculosis H37Rv Rv2559c		Mycobacterium leprae aspS	Mycobacterium tuberculosis H37Rv Rv2575	Saccharomyces cerevisiae \$288C YIR019C sta1	Bacillus subtilis yhgE		Streptomyces coelicolor A3(2) SCE68.13
	db Match	gp:SC1A2_22	gp:AF109162_2		pir.A75169	sp:FHUC_BACSU	pir:D70660	pir.E70660	pir.F70660	sp:SYA_THIFE	sp:Y0A9_MYCTU	-	4 Sp.SYD_MYCLE	sp:Y0BQ_MYCTU	6 SP. AMYH_YEAST	7 Sp:YHGE_BACSU	_	1 gp:SCE68_13
	ORF (bp)	303	1074	909	957	753	828	1167	546	2664	1377	1224	1824	89	2676	1857	648	594
	Terminal (nt)	1721423	1722853	1722202	1723826	1724578	1724612	1725459	1726625	1727385	1730166	1731599	1732988	1735946	1736004	1738713	1740572	1741906
	Initial (nt)	1721725	1721780	1722807		1723826	1725439	1726625	1727170	1730048	1731542	1732822			1738679	1740569	1741219	1741313
	SEO	(a.a.) 5294	5295	5296		5298	5299	5300	5301	5302	5303	5304	5305	5306	5307	5308	5309	
		794 794	795	706		798	799	88	801	802	1803	1804	1805	1806	1807	1808	1809	1810

EP 1 108 790 A2

Table 1 (Confinued)
SEC   Initial   Terminal ORF   Ab Match   Homologous gene   (%)
SEC   Initial   Terminal ORF   db Match   Homologous gene   (%)     1741893   1742606   714   SicE15_13   Streptomyces coelicolor A3(2)   72.8     2313   17428401   1743813   1113   gp. SCE15_13   Streptomyces coelicolor A3(2)   72.8     2314   1743843   1743868   126   Sp. SLFA_PSEAE   Silfaphomyces coelicolor A3(2)   72.8     2315   1747896   1747588   126   Sp. SLFA_PSEAE   Silfaphomyces coelicolor A3(2)   72.8     2316   1747896   1747589   1287   sp. SDHL_ECOLI   Escherichia coli K12 sdaA   46.8     2317   174791   1747980   1287   sp. SDHL_ECOLI   Escherichia coli K12 sdaA   46.8     2318   174791   1747980   1287   sp. SPH_STAAU   Sitaphomyces chrysomalius   35.4     2319   1749963   1749325   639   gp.CJ11168X3_12   Campydobacter jejuni   98.4     2320   175196   1752021   237   2280   gp.AF038651_4   ATCC 13032 orf4     2321   1755496   1752527   342   ATCC 13032 orf4     2322   175548   1755496   156   gp.AF038651_2   ATCC 13032 apt     2323   175548   1755589   150   gp.AF038651_1   ATCC 13032 apt     2324   175548   1755589   150   gp.AF038651_1   ATCC 13032 apt     2325   175728   1757288   1209   sp.SECF_ECOLI   Escherichia coli K12 secf     2326   175728   1757589   120   sp.SECF_ECOLI   Escherichia coli K12 secf     2327   175728   1757289   120   sp.SECF_ECOLI   Escherichia coli K12 secf     2328   175870   1750336   630     2339   175870   1750336   630     2339   175870   1750336   630     2330   175870   1750336   630     2330   175870   1750336   630     2330   175870   1750336   630     2330   175870   1750336   630     2330   175870   1750336   630     2330   175870   1750336   120     2330   175870   1750336   630     2330   175870   1750336   120     2330   175870   1750336   120     2330   175870   1750336   120     2330   175870   1750336   120     2330   175870   1750336   120     2330   175870   1750336   120     2330   175870   1750336   120     2330   175870   1750336   120     2330   175870   1750336   120     2330   175870   1750336   120     2330   175870   1750336   120
Table 1 (continued)   SEQ   Initial   Terminal   CRF   CRF
SEQ Initial Terminal ORF db Match (a.a.) (nt) (nt) (hp) (bp) db Match (a.a.) (nt) (nt) (nt) (hp) db Match (a.a.) 1741893 1742606 714 db Match 5312 1742701 1743813 1113 gp. SCE15_13 5314 1744025 1744519 495 sp. SLFA_PSEAE 5315 1744884 1746230 1347 sp. SDHL_ECOLI 5316 1746728 1747588 861 746233 1686 prt.2423362A 5317 1747918 1746233 1686 prt.2423362A 5318 1749963 1749325 639 gp. CJ11168X3_12 5320 1750427 1750933 507 prt.2313309A 5322 1751497 1752615 2280 gp. AF038651_4 5322 1755486 1752615 2280 gp. AF038651_2 5324 1755489 1752615 2280 gp. AF038651_1 5326 1755748 1755599 150 gp. AF038651_1 5328 1755748 175589 1209 sp. SECF_ECOLI 5329 1759707 1750336 630
SEQ Initial Terminal ORF (nt) (nt) (bp) (aa.) (nt) (nt) (nt) (hp) (5312 1741893 1742606 714 5312 1742701 1743813 1113 5314 1744025 1744519 495 5315 1744984 1746230 1347 5315 1747918 1746233 1686 5316 1749963 1749325 639 5320 1750427 1750933 507 5320 175497 175251 555 532 175549 175251 555 532 175549 1755599 150 5328 1755728 1755789 1755789 175970 1750336 630 5329 175977 1750336 630
SEO (nt) (nt) (nt) (nt) (nt) (nt) (nt) (nt)
SEO (nt) (nt) (aa.) (nt) (5312 1742701 5312 1742701 5313 1743843 5314 1744025 5315 1744884 5315 1749963 5316 1750964 5320 1750964 5322 175186 5326 1755748 5326 1755728 5328 1755728 5328 1755707 5328 1755707 5328 1755707 5329 1759707
SEO NO. (a.a.) 5311 5312 5313 5314 5315 5315 5315 5315 5315 5325 5325 5326 5326 5326 5326 5326 532
<del>                           </del>
NO. (DNA) 1811 1812 1813 1814 1817 1818 1824 1824 1826 1827 1828 1828 1828 1828 1828 1828 1828

EP 1 108 790 A2

		f		_	T	r-	_	Υ			,		1			<del></del>		,	,		
5			Function	protein-export membrane protein	hypothetical protein	holliday junction DNA helicase	holliday junction DNA helicase	crossover junction endodeoxyribonuclease	hypothetical protein	acyl-CoA thiolesterase	hypothetical protein	hypothetical protein	hexosyltransferase or N- acetylglucosaminyl- phosphatidylinositol biosynthetic protein	acyltransferase	CDP-diacylglycerolglycerol-3- phosphate phosphatidyltransferase	histidine triad (HIT) family protein	threonyl-tRNA synthetase	hypothetical protein			
15			Matched length (a.a.)	616	106	331	210	180	250	283	11	170	414	295	82 .	194	647	400			
20			Similarity (%)	52.0	0.99	81.9	74.3	63.3	78.4	68.6	61.3	61.2	49.3	67.8	78.0	78.4	68.9	61.8			
			Identity (%)	24.4	39.6	55.3	45.2	35.6	49.2	38.5	31.5	38.2	21.7	46.4	48.2	54.6	42.0	34.3			
25 30	:	lable 1 (continued)	Homologous gene	Rhodobacter capsulatus secD	Mycobacterium leprae MLCB1259.04	Escherichia coli K12 ruvB	Mycobacterium leprae ruvA	Escherichia coli K12 ruvC	Escherichia coli K12 ORF246 yebC	Escherichia coli K12 tesB	Streptomyces coelicolor A3(2) SC10A5.09c	Mycobacterium tuberculosis H37Rv Rv2609c	Saccharomyces cerevisiae S288C spt14	Streptomyces coelicolor A3(2) SCL2.16c	Mycobacterium tuberculosis H37Rv Rv2612c pgsA	Mycobacterium tuberculosis H37Rv Rv2613c	Bacillus subtilis thrZ	Bacillus subtilis ywbN			
35 40			db Match	prf:2313285A	sp:Y0BD_MYCLE	sp:RUVB_ECOLI	sp:RUVA_MYCLE	sp:RUVC_ECOLI		sp:TESB_ECOLI	gp:SC10A5_9	pir:H70570	sp:GPI3_YEAST	gp:SCL2_16	pir.C70571	pir.D70571	sp:SYT2_BACSU	sp:YWBN_BACSU			
	•		ORF (bp)	1932	363	1080	618	663	753	846	474	462	1083	963	657	099	2058	1206	564	546	735
45			Terminal (nt)	1758803	1761005	1761419	1762517	1763177	1763990	1765015	1766442	1766487	1766948	1768034	1769022	1769681	1770327	1772658	1774444	1773893	1774457
50	·		Initial (nt)	1760734	1761367	1762498	1763134	1763839	1764742	1765860	1765969	1766948	1768030	1768996	1769678	1770340	1772384	1773863	1773881	1774438	1775191
			SEQ NO.	5330	5331	5332	5333	5334	5335	5336	5337	5338	5339	5340	5341	5342	5343	5344	5345	5346	5347
55			SEQ NO (DNA)	1830	1831	1832	1833	1834	1835	1836	1837	1838	1839	1840	1841	1842	1843	1844	1845	1846	1847

EP 1 108 790 A2

	Function	-			·		puromycin N-acetyltransferase										- 1	ferric transport ATP-binding protein					pantothenate metabolism flavoprotein		
	Matched fength (a.a.)						190											202					129		
	Identity Similarity (%)						64.2											28.7					66.7	.,	
İ	Identity (%)			,			36.3											28.7					27.1		
Table 1 (continued)	Homologous gene						Streptomyces anulatus pac											Actinobacillus pleuropneumoniae afuC					Zymomonas mobilis dfp		
	db Match						sp:PUAC_STRLP							·				SP. AFUC_ACTPL					gp:AF088896_20		
•	ORF (bp)	378	594	1407	615	399	567	1086	1101	669	2580	1113	1923	483	189	312	429	597	666	159	1107	420	591	864	420
	Terminal (nt)	1777646	1778037	1778102	1779554	1780507	1781019	1782790	1784381	1783382	1782894	1785732	1786907	1789562	1789768	1790057	1790461	1792438	1793426	1793496	1794820	1795621	1796181	1797049	1797769
	Initial (nt)	1777269	1777444	5350 1779508	1780168	5352 1780905	1781585	1781705	1783281	1784080	1785473	1786844	1788829	1789080	1789580	1789746	1790889	5364 1791842	1792428	1793654	1793714	1795202	1795591	1796186	1797350
	SEQ NO.	5348	5349	5350	5351	5352	5353	5354	5355	5356	5357	5358	5359	5360	5361	5362	5363	5364	5365	5366	5367	5368	5369	5370	5371
	SEQ NO. (DNA)	1848	1849	1850	1851	1852	1853	1854	1855	1856	1857	1858	1859	1860	1861	1862	1863	1864	1865	1866	1867	1868	1869	1870	1871

EP 1 108 790 A2

																					Щ,				
5	Function																			transposon TN21 resolvase			protein-tyrosine phosphatase		
15	Matched length (a.a.)																			186			164		
20	Similarity (%)																			78.0			51.8		
	Identity (%)																			51.1			29.3		
<i>25</i> ਹਿ						}																			
% % Same 1 (continued)	Homologous gene																			Escherichia coli tnpR			Saccharomyces cerevisiae S288C YIR026C yvh1		
40	db Match																			sp:TNP2_ECOLI		Θ	sp.PVH1_YEAST		
	ORF (bp)	120	735	225	894	156	474	753	423	687	429	465	237	681	960	480	681	285	375	612	1005	375	477	726	423
45	Terminal (nt)	1797850	1798023	1799406	1800366	1800449	1801307	1802096	1802155	1803419	1803893	1804598	1804865	1805599	1806686	1807396	1808113	1808421	1808832	1810372	1811545	1811938	1812691	1813606	1812460
50	Initial (nt)	1797969	1798757	1799182	1799473	1800604	1800834	1801344	1802577	1802733	1803465	1804134	1804629	1804919	1805727	1806917	1807433	1808137	1808458	1809761	1810541	1811564	1812215	1812881	1812882
	SEQ NO. (a.a.)	5372	5373	5374	5375	5376	5377	5378	5379	5380	5381	5382	5383	5384	5385	5386	5387	5388	5389	5390	5391	5392	5393	5394	5395
55	SEQ NO.	1872	1873	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	1895

EP 1 108 790 A2

	Γ			T		$\neg$	Т	Т	$\neg$	Т	Т		Т	$\neg$	T		Т			Т			Т	
5		Function	sporulation transcription factor									hypothetical protein					hypothetical protein	insertion element (IS3 related)	insertion element (IS3 related)			single-stranded-DNA-specific exonuclease		primase
<b>15</b>		Matched length (a.a.)	216									545					166	298	101			622		381
20		Similarity (%)	65.7									55.2					75.0	92.6	84.2			50.6		64.3
		Identity (%)	34.3									22.6					63.0	67.9	72.3			24.0		31.8
30	onimined)	is gene	icolor A3(2)									ma MSB8					glutamicum	glutamicum	glutamicum			emi recJ		age phi-01205
30	ומחום ו	Homologous gene	Streptomyces coelicolor A3(2) whiH								•	Thermotoga maritima MSB8 TM1189					Corynebacterium glutamicum	Corynebacterium glutamicum orf2	Corynebacterium glutamicum orf1			Erwinia chrysanthemi recJ		Streptococcus phage phi-O1205 ORF13
35					$\dashv$		-					FF					ပ	ပေခ်	υē					\$0
40		db Match	gp:SCA32WHIH_6								!	pir.C72285					PIR:S60891	pir:S60890	pir.S60889			sp:RECJ_ERWCH		pir:T13302
		ORF (bp)	738	789	456	186	672	417	315	369	207	2202	1746	219	144	429	534	894	294	213	1299	1878	780	1650
45		Terminal (nt)	1814517	1815651	1816128	1816636	1817803	1818219	1818774	1819166	1819748	1820181	1824322	1824589	1824927	1825178	1826557	1825751	1826644	1829688	1832063	1834044	1834149	1838324
50		Initial (nt)	1813780	1814863	1815673	1816451	1817132	1817803	1818460	1818798	1819954	1822382	1822577	1824371	1824784	1825606	1826024	1826644	1826937	1829900	1830765	1832167	1834928	1836675
		SEQ NO.	5396	5397	5398	5399	5400	5401	5402	5403	5404	5405	5406	5407	5408	5409	5410	5411	5412	5413	5414	5415	5416	5417
· 55		SEQ NO. (DNA)	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910	1911	1912	1913	1914	1915	1916	1917

nued)  ne (%) (%) (9%) (9%) (9a3)	
ombe 23.6 52.5 347	oteinase A
nued)  ldentity Similarity (%) (%) (%)  e57 36.7 64.2  combe 28.7 49.8  23.6 52.5	ATP-dependent CIp proteinase ATP-binding subunit
nued) 19	630
25 (eg) (ompe ompe ompe ompe ompe ompe ompe ompe	61.0
25 (ae ATCC are ATCC are Pownbe are ATCC are ATCC	30.2
Table 1 (continued)  Homologous gene Homologous gene Mycoplasma pneumoniae ATCC 29342 yb95 Bacteriophage N15 gene57 Bacteriophage N15 gene57 Schizosaccharomyces pombe SPAPJ760.02c SPAPJ760.02c	Escherichia coli K12 clpA
	Esc
db Match sp:Y018_MYCPN pir.T13144 gp:SPAPJ760_2	sp:CLPA_ECOLI
ORF (bp) 3789 447 534 1839 375 336 528 537 528 537 528 537 528 537 528 536 536 537 538 537 538 537 538 538 538 538 538 538 538 538 538 538	1965
Terminal (nt) 1842681 1842681 1843337 1848474 1849036 1849966 1852440 1852324 1852324 1855237 1855238 1855238 1855238	1860727
1838349 1842235 1842235 1842235 1842804 1845483 18454315 1846698 1847315 1848509 1848988 1848988 1848988 185035 1851220 1851473 1851473 1851473 1851473 1851473	1858763
SEQ NO. (a.a.) 5418 5429 5429 5429 5429 5433 5433 5438 5438 5438 5438 5438 5438	5440
SEQ NO. (DNA.) 1918 1919 1920 1922 1924 1925 1928 1928 1933 1933 1935 1935 1935 1935 1938 1938 1938 1938 1938 1938 1938 1938	1940

EP 1 108 790 A2

5	Function					ATP-dependent helicase					hypothetical protein	deoxynucleotide monophosphate kinase					type II 5-cytosoine methyltransferase	type II restriction endonuclease			hypothetical protein	
15	Matched length (a.a.)					693					224	208					363	358			504	
20	Similarity (%)					45.9					47.8	61.5					99.7	99.7			45.8	
	Identity (%)					21.4					25.9	31.7					99.2	99.7			24.6	
25 (panuj)	ene					us SA20					olor A3(2)	31 gp52					tamicum	tamicum			olor A3(2)	
& Table 1 (continued)	Homologous gene					Staphylococcus aureus SA20 pcrA					Streptomyces coelicolor A3(2) SCH17.07c	Bacteriophage phi-C31 gp52					Corynebacterium glutamicum ATCC 13032 cgllM	Corynebacterium glutamicum ATCC 13032 cgllR			Streptomyces coelicolor A3(2) SC1A2.16c	
40	db Match					sp.PCRA_STAAU					gp:SCH17_7	pri:2514444Y					prf.2403350A	pir.A55225			gp:SC1A2_16	
	ORF (bp)	474	156	324	312	2355	558	378	465	264	777	702	225	2166	273	6507	1089	1074	1521	717	1818	186
45	Terminal (nt)	1861225	1861475	1861519	1862399	1865299	1865822	1866219	1866792	1867095	1867874	1868587	1868671	1868927	1871101	1871380	1879400	1880485	1882470	1884220	1887047	1887590
50	Initial (nt)	1860752	1861320	1861842	1862088	1862945	1865265	1865842	1866328	1866832	1867098	1867886	1868895	1871092	1871373	1877886	1878312	1879412	1883990	1884936	1885230	1887405
	SEQ NO.	5441	5442	5443	5444	5445	5446	5447	5448	-	5450	5451	5452	5453	5454	5455	5456	5457	5458	5459	5460	5461
<i>55</i>	SEQ NO.	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961

EP 1 108 790 A2

			Ţ	T -	T-	$\top$	Τ	T	Τ	Т	T	Τ	Ť	Т	T		Т	T	1	$\top$	T	1	$\top$	$\overline{}$	$\top$	<del></del>
<b>5</b>		Function	SNF2/Rad54 helicase-related protein	otein		otein				endopeptidase Clp ATP-binding							nuclear mitotic apparatus protein									
			SNF2/Rad54 Protein	hypothetical protein		hypothetical protein				endopeptidase							nuclear mitotic									
15		Matched length (a.a.)	06	163		537				724							1004									
20		Similarity (%)	70.0	56.4		47.9				52.5							49.1									
		Identity (%)	46.7	33.1		20.7				25.3							20.1									
25	tinued)	Jene	rans	hi-gle		22-16									ļ											
30	Table 1 (continued)	Homologous gene	Deinococcus radiodurans DR1258	Lactobacillus phage phi-gle Rorf232		Bacillus anthracis pXO2-16				Escherichia coli clpB							Homo sapiens numA									
35				712									-	-	_		エ			_			-	-	-	
40		db Match	gp:AE001973_4	pir.T13226		gp:AF188935_16				sp:CLPB_ECOLI							pir.S23647									
		ORF (bp)	351	864	330	1680	1206	1293	2493	1785	621	1113	846	981	879	198	2766	009	1251	969	714	1008	1659	1488	399	1509
45		Terminal (nt)	1887688	1888231	1889859	1890028	1891832	1893388	1894739	1897374	1899233	1899804	1901066	1902955	1902005	1903225	1903113	1905973	1906664	1907965	1908785	1909501	1910642	1912333	1913973	1914725
50		Initial (nt)	1888038	1889094	1889530	1891707	1893037	1894680	1897231	1899158	1899853	1900916	1901911	1901975	1902883	1903028	1905878	1906572	1907914	1908660	1909498	1910508	1912300	1913820	1914371	1916233
		SEQ NO.	5462	5463	5464	5465	5466	5467	5468	5469	5470	5471	5472	5473	5474	5475	5476	5477	5478	5479	5480	5481	5482	5483		5485
55	į	SEQ NO.	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983		1985

EP 1 108 790 A2

			_			$\overline{}$	_	_		_			_	_								_				
5		ے													a											
		Function										submaxillary apomucin			modification methylase					otein			otein			
10												dlary a			ation m			ļ		tical pr			tical pro			
												submax			modific					hypothetical protein			hypothetical protein			
15		Matched length (a.a.)										1408	$\top$		61					114			328			
20		Similarity (%)										49.2			65.6					58.8			54.6			
		Identity (%)										23.2			42.6					38.6			27.1			
25	ned)	<u> </u>																		Sis			:=			
	Table 1 (continued)	Homologous gene										stica			ecoR1		į			Mycobacterium tuberculosis H37Rv Rv1956			Methanococcus jannaschii MJ0137			
30	Table 1	fomolog										fa dome			hia coli					terium t			coccus		ļ 	
	·											Sus scrofa domestica			Escherichia coli ecoR1					lycobac 137Rv R			lethano J0137			
35				-				-				0)	<del> </del>							≥I						
		db Match	!									660			1_ECOLI					638			sp:Y137_METJA			
40	:											pir.T03099			sp:MTE1					pir:H70638			sp:Y13			
		ORF (bp)	360	222	312	645	759	549	930	306	357	4464	579	945	171	375	1821	201	468	381	507	837	942	624	210	534
45		Terminal (nt)	1916733	1917165	1917329	1917564	1918703	1919646	1920347	1925695	1926038	1921547	1926259	1927245	1928381	1928908	1929059	1930990	1931421	1931935	1932373	1933522	1934971	1936849	1937411	1937486
50		Initial (nt)	1916374	1916944	1917640	1918208	1919461	1920194	1921276	1925390	1925682	1926010	1926837	1928189	1928211	1928534	1930879	1931190	1931888	1932315	1932879	1934358	1935912	1936226	1937202	1938019
	ļ	SEQ NO.	5486	5487	5488	5489	5490	5491	5492	5493	5494	5495	5496	5497	5498	5499	5500	5501	5502	5503	5504	5205	5506	5507	5508	5509
55		SEQ NO.	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009

EP 1 108 790 A2

			,				- 1					_			$\neg$			$\overline{}$		_				_			
5			Function										surface protein				major secreted protein PS1 protein precursor			DNA topoisomerase III					major secreted protein PS1 protein precursor		
15			Matched length (a.a.)										304				270			597					344		
20			Similarity (%)										44.1				54.4			50.9					54.7		
			Identity (%)										23.0				30.7			23.8					29.7		
25	<del>1</del>	lable 1 (continued)	ns gene										alis esp	-			glutamicum ivum) ATCC			88					lutamicum vum) ATCC		
30	H	lable 1 (	Homologous gene										Enterococcus faecalis esp				Corynebacterium glutamicum (Brevibacterium flavum) ATCC 17965 csp1			Escherichia coli topB					Corynebacterium glutamicum (Brevibacterium flavum) ATCC 17965 csp1		
<i>35</i>			db Match										prf.2509434A				sp:CSP1_CORGL			sp:TOP3_ECOLI					sp:CSP1_CORGL (		
			ORF (bp)	1191	534	588	444	753	303	216	309	885	828	297	381	429	1581	2430	867	2277	2085	891	432	744	1887	291	
45			Terminal (nt)	1940135	1938531	1940844	1941550	1941732	1942812	1943310	1943653	1944564	1944608	1945595	1945952	1946609	1947070	1949021	1951619	1952546	1956203	1958450	1959765	1960371	1961114	1963139	
50			Initial (nt)	1938945	1939064	1940257	1941107	1942484	1942510	5516 1943095	1943345	1943680	5519 1945435	1945891	1946332	1947037	1948650	1951450	1952485	1954822	1958287	1959340	1960196	1961114	1963000	5532 1963429	
			SEQ NO. (a.a.)	5510	5511	5512	5513	5514	5515		5517	5518		5520	5521	5522		5524		5526	5527	5528	5529	5530	5531	5532	
<i>55</i>			SEQ NO (DNA)	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020		2022	2023	2024		2026	2027	2028	2029	2030	2031	2032	

5	Function								*						single stranded DNA-binding protein							*					
10	E P				thermonuclease										single stranded DN								serine protease				
15	Matched length (a.a.)				227										225								249				
20	Similarity (%)				57.7										59.1								52.6				
	Identity (%)				30.4										24.9								25.7				
25 (penci)	s gene				reus nuc																		AgSP24D				
7able 1 (continued)	Homologous gene				Staphylococcus aureus nuc										Shewanella sp. ssb		-						Anopheles gambiae AgSP24D				
<i>40</i>	db Match				sp:NUC_STAAU										prf:2313347B								sp.S24D_ANOGA A				
	ORF (bp)	1230	1176	357	684 sp	147	564	1452	459	1221	1419	591	396	237	624 prf	579	462	202	588	333	558	570	912 sp.	693	366	747	180
45	Terminal (nt)	1963514	1964727	1965911	1966984	1967289	1968167	1969715	1970203	1971474	1973090	1973737	1974204	1974503	1975794	1976494	1976983	1977549	1978329	1978721	1979217	1979808	1980885	1981657	1982028	1982817	1981912
50	Initial (nt)	1964743	1965902	1966267	1966301	1967435	1967604	1968264	1969745	1970254	1971672	1973147	1973809	1974267	1975171	1975916	1976522	1977043	1977742	1978389	1978660	1979239	1979974	1980965	1981663	1982071	1982091
	SEO NO.	5533	5534	5535	5536	5537	5538	5539	5540	5541	5542	5543	5544	5545	5546	5547	5548	5549	5550	5551	5552	5553	5554	5555	5556	:	5558
55	SEQ NO (DNA)	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058

				-			. —														
5	Function								integrase	transposase (divided)	transposase (divided)		transposition repressor	insertion element (IS3 related)	transposase					major secreted protein PS1 protein precursor	integrase
15	Matched length (a.a.)								40e	124 t	117 (		31	43	270 t		-			153 F	223
20	Similarity (%)								55.9	94.4	84.6		8.96	88.4	53.7					37.0	56.1
	Identity (%)								29.6	83.9	6.07		80.7	74.4	31.1					25.0	28.7
Table 1 (continued)	Homologous gene								phage L5 int	actofermentum	actofermentum		actofermentum	n glutamicum	elicolor A3(2)					glutamicum flavum) ATCC	shage L5 int
Table 1	Нотово					:			Mycobacterium phage L5 int	Brevibacterium lactofermentum CGL 2005 ISaB1	Brevibacterium lactofermentum CGL2005 ISaB1		Brevibacterium lactofermentum CGL2005 ISaB1	Corynebacterium glutamicum orf 1	Streptomyces coelicolor A3(2) SCJ11.12					Corynebacterium glutamicum (Brevibacterium flavum) ATCC 17965 csp1	Mycobacterium phage L5 int
40	db Match								Sp:VINT_BPML5	gsp:R23011	gsp:R23011		gsp:R21601	pir.S60889	gp:SCJ11_12					sp:CSP1_CORGL	sp:VINT_BPML5
	ORF (bp)	363	273	264	234	342	273	303	1149 sp	390 gs	417 gs	207	114 gs	135 pii	828 gp	354	891	432	744	1584 sp	687 sp
45	Terminal (nt)	1983548	1983883	1984181	1984450	1984728	1985364	1985071	1985442 1	1987507	1987887	1988589	1988370	1988530	1988778	1991020	1989874	1991189	1991795	1992538 1	1994608
50	Initial (nt)	1983186	1983611	1983918	1984217	1984387	1985092	1985373	1986590	1987896	1988303	1988383	1988483	1988664	1989605	1990667	1990764	1991620	1992538	1994121	1995294
	SEQ NO.	5559	2560	5561	2955	5563	5564	5955	5566	5567	5568	5569	5570	5571	5572	5223	5574	525	5576	5577	5578
55	SEQ NO (DNA)	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078

	· .					•		Ţ					<b>a</b>						
5	c	ansporter				s protein	orotein	reductase					5-phosphati	se			osphate		
10	Function	sodium-dependent transporter	hypothetical protein			riboflavin biosynthesis protein	potential membrane protein	methionine sulfoxide reductase		hypothetical protein	hypothetical protein	ribonuclease D	1-deoxy-D-xylulose-5-phosphate synthase	RNA methyltransferase		hypothetical protein	deoxyuridine 5'-triphosphate nucleotidohydrolase	hypothetical protein	
15	Matched length (a.a.)	88	92			233	384	126		232	201	371	618	472		268	140	150	
20	Similarity (%)	76.1	81.5			64.4	71.9	67.5		77.2	78.6	52.8	78.5	52.3		62.7	82.1	7.07	
•	Identity (%)	39.8	48.9			33.5	42.5	41.3		55.2	55.7	25.9	55.3	25.4		38.1	55.0	46.0	
25 (juned)	зепе	9699				culosis	culosis	nii msrA		culosis	culosis	zae Rd	.190 dxs	a MSB8		culosis	olor A3(2)	rculosis	
8 Table 1 (continued)	Homologous gene	Helicobacter pylori 26695 HP0214	Bacillus subtilis yxaA			Mycobacterium tuberculosis H37Rv Rv2671 ribD	Mycobacterium tuberculosis H37Rv Rv2673	Streptococcus gordonii msrA		Mycobacterium tuberculosis H37Rv Rv2676c	Mycobacterium tuberculosis H37Rv Rv2680	Haemophilus influenzae KW20 H10390 rnd	Streptomyces sp. CL190 dxs	Thermotoga maritima MSBB TM1094		Mycobacterium tuberculosis H37Rv Rv2696c	Streptomyces coelicolor A3(2) SC2E9.09 dut	Mycobacterium tuberculosis H37Rv Rv2698	
<b>35</b>								2					-				o.		
40	db Match	pir.F64546	sp:YXAA_BACSU			pir.C70968	pir.E70968	gp:AF128264_		pir:H70968	pir:C70528	sp:RND_HAEIN	gp:AB026631_1	pir.E72298		pir.C70530	sp:DUT_STRCO	pir.E70530	
,	ORF (bp)	306	432	345	336	969	1254	408	426	969	624	1263	1908	1236	282	861	447	549	207
45 ·	Terminal (nt)	1995783	1996537	1997112	1997503	1998240	1999542	1999949	1999707	2000521	2002112	2003334	2003402	2005462	2006979	2006777	2007738	2008798	2008876
50	· Initial	1996088	1996106	1996768	1997168	1997545	1998289	1999542	2000132	2001216	2001489	2002072	2005309	2006697	2006698	2007637	2008184	2008250	2009082
	SEQ NO.	5579	5580	5581	5582	5583	5584	5585	5586	5587	5588	5589	2590	5591	5592	5593	5594	5595	5596
55	SEQ	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096

5	uo		or protein	kinase	polymerase	ine protein			ine protein				ssor or	protein	erase			helicase
10	Function	hypothetical protein	extragenic suppressor protein	polyphosphate glucokinase	sigma factor or RNA polymerase transcription factor	hypothetical membrane protein		hypothetical protein	hypothetical membrane protein	hypothetical protein	transferase	hypothetical protein	iron dependent repressor or diphtheria toxin repressor	putative sporulation protein	UDP-glucose 4-epimerase		hypothetical protein	ATP-dependent RNA helicase
15	Matched length (a.a.)	100	198	248	200	422		578	127	9/	523	144	228	77	329		305	661
20	Similarity (%)	81.0	68.2	80.2	98.6	51.4		80.8	59.1	85.5	61.2	100.0	9.66	64.0	99.1		79.0	50.7
	identity (%)	58.0	38.4	54.4	98.0	23.9		61.3	32.3	65.8	33.5	97.2	98.7	62.0	99.1		45.3	24.4
<i>25</i>		sis	_ 	sis	cum			sis	sis	Sis	43(2)	cum	cum	ns	cum		sis	e
30 alder	ous gen	uberculo	K12 suh	uberculo ppgK	n glutam	/rkO		uberculo	uberculo	uberculo	elicolor,	ո glutami የF1	ր glutami R	reofacie	n glutarni evibacte galE		uberculo	cerevisia
10 H	Homologous gene	Mycobacterium tuberculosis H37Rv Rv2699c	Escherichia coli K12 suhB	Mycobacterium tuberculosis H37Rv RV2702 ppgK	Corynebacterium glutamicum sigA	Bacillus subtilis yrkO		Mycobacterium tuberculosis H37Rv Rv2917	Mycobacterium tuberculosis H37Rv Rv2709	Mycobacterium tuberculosis H37Rv Rv2708c	Streptomyces coelicolor A3(2) SCH5.08c	Corynebacterium glutamicum ATCC 13869 ORF1	Corynebacterium glutamicum ATCC 13869 dtxR	Streptomyces aureofaciens	Corynebacterium glutamicum ATCC 13869 (Brevibacterium lactofermentum) galE		Mycobacterium tuberculosis H37Rv Rv2714	Saccharomyces cerevisiae YJL050W dob1
35					<b>.</b>				~ +	<u> </u>	0) 0)	0 4	OW				· ·	
40	db Match	pir.F70530	SP. SUHB_ECOLI	sp.PPGK_MYCTU	prf.2204286A	Sp.YRKO_BACSU		sp:Y065_MYCTU	pir:H70531	pir:G70531	gp:SCH5_8	prf:2204286C	pir:140339	GP:AF010134	Sp:GALE_BRELA		pir.E70532	sp:MTR4_YEAST
	ORF (bp)	291	816	828	1494	1335	537	1710	636	237	1533	432	684	234	987	1323	957	2550
45	Terminal (nt)	2009280	2009724	2011382	2013356	2014162	2015585	2016257	2018754	2017966	2020276	2020724	2022949	2022313	2023945	2023948	2026379	2029043
50	Initial (nt)	2009570	2010539	2010555	2011863	2015496	2016121	2017966	2018119	2018202	2018744	2020293	2022266	2022546	2022959	2025270	2025423	2026494
	SEQ NO.	5597	5598	5599	2600	5601	5602	5603	5604	5605	9099	5607	5608	5609	5610	5611	5612	5613
55	SEQ NO. (DNA)	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

tinued)
Table 1 (contin

			,														_	_		
	Function	hydrogen peroxide-inducible genes activator		ATP-dependent helicase	regulatory protein		SOS regulatory protein	galactitol utilization operon repressor	phosphofructokinase (fructose 1- phosphate kinase)	phosphoenolpyruvate-protein phosphotransferase	glycerol-3-phosphate regulon repressor	1-phosphofructokinase or 6- phosphofructokinase	PTS system, fructose-specific IIBC component	phosphocarrier protein		uracil permease	ATP/GTP-binding protein			diaminopimelate epimerase
	Matched length (a.a.)	299		1298	145		222	245	320	592	262	345	549	81		407	419			269
	Similarity (%)	65.6		76.2	86.2		71.6	8.79	55.6	64.0	62.6	55.7	9.69	71.6		70.5	80.0			64.7
	Identity (%)	35.8		49.2	61.4		46.9	33.9	27.2	34.3	26.7	33.0	43.0	37.0		39.1	54.4			33.5
ומסור ו (בפונוווומבת)	Homologous gene	Escherichia coli oxyR		Escherichia coli hrpA	Streptomyces clavuligerus nrdR		Bacillus subtilis dinR	Escherichia coli K12 gatR	Streptomyces coelicolor A3(2) SCE22.14c	Bacillus stearothermophilus ptsl	Escherichia coli K12 glpR	Rhodobacter capsulatus fruK	Escherichia coli K12 fruA	Bacillus stearothermophilus XL- 65-6 ptsH		Bacillus caldolyticus pyrP	Streptomyces fradiae orf11*			Haemophilus influenzae Rd KW20 Hi0750 dapF
	db Match	sp:OXYR_ECOLI		sp:HRPA_ECOLI	gp:SCAJ4870_3		sp:LEXA_BACSU	sp:GATR_ECOLI	gp:SCE22_14	sp:PT1_BACST	sp:GLPR_ECOLI	sp:K1PF_RHOCA	sp:PTFB_ECOLI	sp:PTHP_BACST		sp:PYRP_BACCL	gp:AF145049_8			sp:DAPF_HAEIN
	ORF (bp)	981	1089	3906	450	420	969	777	096	1704	792	066	1836	267	582	1287	1458	786	537	831
	Terminat (nt)	2030157	2030277	2035383	2035431	2035990	2037507	2038591	2039550	2039618	2042519	2043508	2045571	2046028	2046714	2047320	2048650	2051106	2051842	2051845
	Initial (nt)	2029177	2031365	2031478	2035880	5618 2036409	5619 2036812	2037815	2038591	2041321	2041728	5624 2042519	2043736	2045762	2047295	2048606	2050107	2050321	2051306	2052675
	SEQ NO. (a.a.)	5614	5615	5616	5617	5618	5619	5620	5621	5622	5623	5624	5625	5626	5627	5628	5629	5630	5631	5632
	SEQ NO.	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132

				_												_			
5		u	phate					ne protein		TP-binding	es predicted to r vaccines and	ystem	ystem					ATP-binding	ne protein
10		Function	tRNA delta-2- isopentenylpyrophosphate transferase		hypothetical protein			hypothetical membrane protein	hypothefical protein	glutamate transport ATP-binding protein	Neisserial polypeptides predicted to be useful antigens for vaccines and diagnostics	glutamate transport system permease protein	glutamate transport system permease protein	regulatory protein	hypothetical protein		biotin synthase	putrescine transport ATP-binding protein	hypothetical membrane protein
15 <sup>·</sup>		Matched length (a.a.)	300		445			190	494	242	7.1	225	273	142	29		197	223	228
20		Similarity (%)	68.7		7.57			63.7	86.4	9.66	73.0	100.0	9.66	6.99	71.6		61.4	69.5	58.8
		Identity (%)	40.0		48.5		,	29.0	68.4	93.6	66.0	100.0	99.3	34.5	40.3		33.0	33.2	24.6
25	ontinued)	s gene	2 miaA		erculosis			erculosis	rae	lutamicum	eae	lutamicum	lutamicum vum) ATCC	rae recX	erculosis		bioY	2 potG	u_
	Table 1 (continued)	Homologous gene	Escherichia coli K12 miaA		Mycobacterium tuberculosis H37Rv Rv2731			Mycobacterium tuberculosis H37Rv Rv2732c	Mycobacterium leprae B2235_C2_195	Corynebacterium glutamicum ATCC 13032 gluA	Neisseria gonorrhoeae	Corynebacterium glutamicum ATCC 13032 gluC	Corynebacterium glutamicum (Brevibacterium flavum) ATCC 13032 gluD	Mycobacterium leprae recX	Mycobacterium tuberculosis H37Rv Rv2738c		Bacillus sphaericus bioY	Escherichia coli K12 potG	Bacillus subtilis ybaF
35					ΣI			ΣI			Z				ΣI				ă
40		db Match	sp:MIAA_ECOLI		pir.B70506			pir.C70506	sp:Y195_MYCLE	sp:GLUA_CORGL	GSP:Y75358	sp:GLUC_CORGL	sp:GLUD_CORGL	sp:RECX_MYCLE	pir:A70878		sp:BIOY_BACSH	sp:POTG_ECOLI	pir.F69742
٠	!	ORF (bp)	903	675	1359	1020	1023	699	1566	726	219	684	819	597	234	738	576	669	609
45		Terminal (nt)	2052684	2053609	2055761	2054724	2056787	2057120	2057855	2060499	2060196	2062312	2063259	2063298	2065394	2065667	2067141	2067866	2068474
50		Initial (nt)	2053586	2054283	2054403	2055743	2055765	2057788	2059420	2059774	2060414	2061629	2062441	2063894	2065627	2066404	2066566	2067168	2067866
	i	SEQ NO. (a.a.)	5633	5634	5635	5636	5637	5638	5639	5640	5641	5642	5643	5644	5645	5646	5647	5648	5649
<i>55</i>		SEQ NO. (DNA)	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149

5		Function	notein	hypothetical protein (35kD protein)	regulator (DNA-binding protein)	competence damage induced proteins	phosphotidylglycerophosphate synthase	oratein	surface protein (Peumococcal surface protein A)		ance protein
10			hypothetical protein	hypothetical p	regulator (DN	competence of proteins	phosphotidylg synthase	hypothetical protein	surface protein (P surface protein A)		tellurite resistance protein
15		Matched length (a.a.)	228	269	83	165	160	117	30		358
20		Identity Similarity (%)	78.5	89.6	78.3	68.5	72.5	52.1	70.0		59.8
		Identity (%)	41.7	72.5	54.2	41.8	38.8	24.8	0.09		31.0
25	ontinued)	gene	erculosis	erculosis	erculosis	ımoniae R6X	lenes pgsA	6	ımoniae		0
<i>30</i>	Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis	Mycobacterium tuberculosis H37Rv RV2744C	Mycobacterium tuberculosis H37Rv Rv2745c	Streptococcus pneumoniae R6X cinA	Streptococcus pyogenes pgsA	Arabidopsis thaliana ATSP:T16118.20	Streptococcus pneumoniae DBL5 pspA		Escherichia coli terC
40		db Match	pir:B60176	sp:35KD_MYCTU	pir:H70878	sp.CINA_STRPN	prf.2421334D	pir.T10688	gp:AF071810_1		nrf 2119295D
		ORF (bp)	069	828	321	516	603	285	117	813	1107
45		Terminal (nt)	2069392	2068556	2069616	2069997	2070519	2071599	2071740	2072878	2072905 2071799
50		Initial (nt)	2068703	2069383	2069936	2070512	2071121	2071315	2071624	2072066	
		SEQ NO (a.a.)	5650	5651	5652	5653	5654	5655	5656	5657	565R
		008	18	-	22	1 23	1 7	52	ဖွ	15	ď

_																			_
	Function	hypothetical protein	hypothetical protein (35kD protein)	regulator (DNA-binding protein)	competence damage induced proteins	phosphotidylglycerophosphate synthase	hypothetical protein	surface protein (Peumococcal surface protein A)		tellurite resistance protein	stage III sporulation protein E	hypothetical protein	hypothetical protein	hypothetical protein	•		guanosine pentaphosphate synthetase	30S ribosomal protein S15	nucleoside hydrolase
	Matched length (a.a.)	228	269	83	165	160	117	30		358	845	216	645	250			742	89	319
	Similarity (%)	78.5	89.6	78.3	68.5	72.5	52.1	70.0		59.8	64.6	61.0	99.4	93.6			85.3	88.8	63.3
	Identity (%)	41.7	72.5	54.2	41.8	38.8	24.8	0.09		31.0	38.0	33.3	99.1	99.2			65.4	64.0	35.1
	Homologous gene	Mycobacterium tuberculosis	Mycobacterium tuberculosis H37Rv RV2744C	Mycobacterium tuberculosis H37Rv Rv2745c	Streptococcus pneumoniae R6X cinA	Streptococcus pyogenes pgsA	Arabidopsis thaliana ATSP:T16118.20	Streptococcus pneumoniae DBL5 pspA		Escherichia coli terC	Bacillus subtilis 168 spolllE	Streptomyces coelicolor A3(2) SC4G6.14	Corynebacterium glutamicum ATCC 13032 orf4	Corynebacterium glutamicum (Brevibacterium lactofermentum) ATCC 13869 orf2			Streptomyces antibioticus gpsl	Bacillus subtilis rpsO	Leishmania major
	db Match	pir.B60176	sp:35KD_MYCTU	pir.H70878	sp.CINA_STRPN	prf.2421334D	pir.T10688	gp:AF071810_1		prf.2119295D	sp:SP3E_BACSU	gp:SC4G6_14	sp:YOR4_CORGL	sp:YDAP_BRELA			prf.2217311A	pir.F69700	prf:2518365A
	ORF (bp)	969	828	321	516	603	285	117	813	1107	2763	633	2154	750	669	264	2259	267	948
	Terminal (nt)	2069392	2068556	2069616	2069997	2070519	2071599	2071740	2072878	2071799	2073294	2076392	2077122	2080387	2082813	2082105	2082932	2085436	2085879
	Initial (nt)	2068703	2069383	2069936	2070512	2071121	2071315	2071624	2072066	2072905	2076056	2077024	2079275	2081136	2082115	2082368	2085190	2085702	2086826
	SEQ NO.	5650	5651	5652	5653	5654	5655	5656	5657	5658	5659	2660	5661	5662	5663	5664	5665	5666	2995
	SEQ NO. (DNA)	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167

peptidetransport system ABC-transporter ATP-binding protein

oligopeptide permease

292 552

69.2 81.3

57.6 38.4

Mycobacterium tuberculosis H37Rv Rv3663c dppD

Bacillus subtilis spo0KC

prf: 1709239C

666

2103973

5683 2102975

2183 2184

5684

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

	_																
5	-	Function	bifunctional protein (riboflavin kinase and FAD synthetase)	tRNA pseudouridine synthase B	protein	protein	erase	DNA damaged inducible protein f	protein	ribosome-binding factor A	translation initiation factor IF-2	protein	n-utilization substance protein (transcriptional termination/antitermination factor)		protein	ling protein	peptidetransport system permease
10			bifunctional protein (n and FAD synthetase)	tRNA pseud	hypothetical protein	hypothetical protein	phosphoesterase	DNA damag	hypothetical protein	ribosome-bi	translation in	hypothetical protein	n-utilization sul (transcriptional termination/ant		hypothetical protein	peptide-binding protein	peptidetrans
15		Matched length (a.a.)	329	303	47	237	273	433	308	108	1103	83	352		165	534	337
20		Similarity (%)	0.67	61.7	73.0	62.5	6.89	78.8	70.8	70.4	62.9	66.3	71.0		65.5	6.09	69.4
		Identity (%)	56.2	32.7	65.0	42.2	46.9	51.0	36.7	32.4	37.7	44.6	42.3		34.6	25.3	37.7
<i>25 30</i>	Table 1 (continued)	Homologous gene	Corynebacterium ammoniagenes ATCC 6872 ribF	Bacillus subtilis 168 truB	sterium enes	Streptomyces coelicolor A3(2) SC5A7.23	Mycobacterium tuberculosis H37Rv Rv2795c	Mycobacterium tuberculosis H37Rv Rv2836c dinF	Mycobacterium tuberculosis H37Rv Rv2837c	Bacillus subtilis 168 rbfA	Stigmatella aurantiaca DW4 infB	Streptomyces coelicolor A3(2) SC5H4.29	Bacillus subtilis 168 nusA		Mycobacterium tuberculosis H37Rv Rv2842c	Bacillus subtilis 168 dppE	Escherichia coli K12 dppB
	Ē	Ť	Corynebacterium ammoniagenes A	Bacillus su	Corynebacterium ammoniagenes	Streptomy SC5A7.23	Mycobacterium t H37Rv Rv2795c	Mycobacte H37Rv Rv	Mycobacterium to H37Rv Rv2837c	Bacillus st	Stigmatell	Streptomy SC5H4.29	Bacillus su		Mycobacto H37Rv Rv	Bacillus st	Escherich
<i>35</i> <i>40</i>		db Match	sp:RIBF_CORAM	sp.TRUB_BACSU	PIR:PC4007	gp:SC5A7_23	pir.B70885	pir:G70693	pir.H70693	sp:RBFA_BACSU	sp:IF2_STIAU	gp:SC5H4_29	sp:NUSA_BACSU		pir.E70588	sp:DPPE_BACSU	sp:DPPB_ECOLI
		ORF (bp)	1023	891	228	651	804	1305	966	447	3012	336	966	1254	534	1602	924
45		Terminal (nt)	2086919	2088863	2087954	2089218	2089861	2090751	2092051	2093055	2093712	2096844	2097380	2099815	2098412	2101841	2102946
50		Initial (nt)	2087941	2087973	2088181	2089868	2090664	2092055	2093046	2093501	2096723	2097179	2098375	2098562	2098945	2100240	2102023
		SEQ NO.	5668	5669	5670	5671	5672	5673	5674	5675	5676	5677	5678	5679	5680	5681	5682
55		SEQ NO.	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182

5		Function	prolyl-tRNA synthetase	hypothetical protein	magnesium-chelatase subunit	magnesium-chelatase subunit	uroporphyrinogen III methyltransferase	hypothetical protein	hypothetical protein	hypothetical protein	glutathione reductase					methionine aminopeptidase	penicillin binding protein	response regulator (two-component system response regulator)	two-component system sensor histidine kinase	hypothetical membrane protein
		·	prolyl-tRI	hypothet	magnesi	magnesi	uroporph methyltra	hypothet	hypothet	hypothet	glutathio					methioni	penicillin	system r	two-component histidine kinase	hypothe
15		Matched length (a.a.)	578	243	37	342	237	488	151	338	466					252	630	216	424	360
20		Similarity (%)	84.6	65.0	60.7	9.69	73.8	68.7	62.3	65.7	76.6					75.8	56.5	72.2	56.8	58.1
		Identity (%)	67.0	39.5	32.4	46.5	49.0	41.2	35.1	37.6	53.0					47.2	27.3	44.0	29.5	24.4
25	(pant	ne	losis	r A3(2)	les ATCC	h	denreichii	NCIB	r A3(2)	ılosis	.C1100					ар	erus pcbR	heriae	heriae	ıns
30	Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv2845c proS	Streptomyces coelicolor A3(2) SCC30.05	Rhodobacter sphaeroides ATCC 17023 bchD	Heliobacillus mobilis bchl	Propionibacterium freudenreichii cobA	Clostridium perfringens NCIB 10662 ORF2	Streptomyces coelicolor A3(2) SC5H1 10c	Mycobacterium tuberculosis H37Rv Rv2854	Burkholderia cepacia AC1100 gor					Escherichia coli K12 map	Streptomyces clavuligerus pcbR	Corynebacterium diphtheriae chrA	Corynebacterium diphtheriae chrS	Deinococcus radiodurans DRA0279
35				ळ ळ		Ĭ	<u> </u>		क ल	ΣÏ							S	ਹ ਰ	ਹ ਹ	
40		db Match	sp:SYP_MYCTU	gp:SCC30_5	sp BCHD_RHOSH	prf.2503462AA	prf:2108318B	sp:YPLC_CLOPE	gp:SC5H1_10	pir.A70590	sp:GSHR_BURCE					SP:AMPM_ECOLI	prf. 2224268A	prf:2518330B	prf.2518330A	gp: AE001863_70
		ORF (bp)	1764	735	759	1101	750	1422	006	1014	1395	942	474	357	729	789	1866	630	1149	957
45		Terminal (nt)	2105801	2108386	2108389	2109155	2110434	2112659	2112717	2116774	2118310	2117015	2119080	2119495	2120356	2120359	2121296	2123219	2123848	2126045
50		Initial (nt)	2107564	2107652	2109147	2110255	2111183	2111238	2113616	2115761	2116916	2117956	2118607	2119139	2119628	2121147	2123161	2123848	2124996	2125089
		SEQ NO.	5685	5686	5687	5688	5689	2690	5691	2695	5693	5694	5695	5696	5697	5698	5699	5700	5701	5702
<i>5</i> 5		SEQ NO.	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202

5		Function	ABC transporter		hypothetical protein (gcpE protein)		hypothetical membrane protein	polypeptides can be used as vaccines against Chlamydia trachomatis	1-deoxy-D-xylulose-5-phosphate reductoisomerase				ABC transporter ATP-binding protein	pyruvate formate-lyase 1 activating enzyme	hypothetical membrane protein	phosphatidate cytidylyltransferase	ribosome recycling factor	uridylate kinase		elongation factor Ts	30S ribosomal protein S2
15		Matched length (a.a.)	225		359		405 1	147	312				245	356	94	294 p	185 r	109		280	254
20		Similarity (%)	71.1		73.8		73.6	43.0	42.0				75.1	78.0	74.5	56.5	84.3	43.1		76.8	83.5
		Identity (%)	37.3		44.3		43.0	36.0	22.8				37.1	0.99	41.5	33.3	47.0	28.4		49.6	54.7
25	cg)						sis	-					188	sis	sis	eo .		а ругН		43(2)	
30	lable i (confinueu)	Homologous gene	Bacillus subtilis 168 yvrO		Escherichia coli K12 gcpE		Mycobacterium tuberculosis H37Rv Rv2869c	Chlamydia trachomatis	Escherichia coli K12 dxr				Thermotoga maritima MSB8 TM0793	Mycobacterium tuberculosis H37Rv	Mycobacterium tuberculosis H37Rv Rv3760	Pseudomonas aeruginosa ATCC 15692 cdsA	Bacillus subtilis 168 frr	Pseudomonas aeruginosa pyrH		Streptomyces coelicolor A3(2) SC2E1.42 tsf	Bacillus subtilis rpsB
40		db Match	prf:2420410P		sp:GCPE_ECOL!		pir.G70886	GSP:Y37145	sp:DXR_ECOLI				pir:B72334	sp:YS80_MYCTU	pir.A70801	sp:CDSA_PSEAE	sp.RRF_BACSU	prf:2510355C		sp.EFTS_STRCO	pir:A69699
		ORF (bp)	069	162	1134	612	1212	645	1176	441	480	1578	855	1098	258	855	555	729	861	825	816
45		Terminal (nt)	2126753	2126926	2127350	2129461	2128669	2130950	2129903	2131762	2131247	2131825	2133406	2134454	2136141	2136235	2137286	2137936	2139854	2139003	2140071
50		Initial (nt)	2126064	2127087	2128483	2128850	2129880	2130306	2131078	2131322	2131726	2133402	2134260	2135551	2135884	2137089	2137840	2138664	2138994	2139827	2140886
		SEQ NO. (a.a.)	5703	5704	5705	5706	5702	5708	5709	5710	5711	5712	5713	5714	5715	5716	5717	5718	5719	5720	5721
55		SEQ NO (DNA)	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

	Function	hypothetical protein	site-specific recombinase	hypothetical protein	Mg(2+) chelatase family protein	hypothetical protein	hypothetical protein	ribonuclease HII		signal peptidase	Fe-regulated protein		50S ribosomal protein L19	thiamine phosphate pyrophosphorylase	oxidoreductase	thiamine biosynthetic enzyme thiS (thiG1) protein	thiamine biosynthetic enzyme thiG protein	molybdopterin biosynthesis protein
	Matched length (a.a.)	120	297	395	504	119	101	190		285	323		111	225	376	62	251	437
	Similarity (%)	58.0	68.7	66.8	75.8	72.3	0.96	5.69		61.1	59.1		88.3	6.09	64.1	74.2	76.9	56.8
	Identity (%)	46.0	40.1	39.8	46.6	40.3	68.3	42.6		32.3	25.4		70.3	28.4	34.0	37.1	48.2	30.2
Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv2891	Proteus mirabilis xerD	Mycobacterium tuberculosis H37Rv Rv2896c	Mycobacterium tuberculosis H37Rv Rv2897c	Mycobacterium tuberculosis H37Rv Rv2898c	Mycobacterium tuberculosis H37Rv Rv2901c	Haemophilus influenzae Rd H11059 rnhB		Streptomyces lividans TK21 sipY	Staphylococcus aureus sirA		Bacillus stearothermophilus rplS	Bacillus subtilis 168 thiE	Streptomyces coelicolor A3(2) SC6E10.01	Escherichia coli K12 thiS	Escherichia coli K12 thiG	Emericella nidulans cnxF
	db Match	sp:YS91_MYCTU	prf:2417318A .	sp:YX27_MYCTU	sp:YX28_MYCTU	sp:YX29_MYCTU	sp:YT01_MYCTU	sp:RNH2_HAEIN		prf.2514288H	prf.2510361A	i	sp:RL19_BACST	sp:THIE_BACSU	gp:SC6E10_1	sp:THIS_ECOLI	sp:THIG_ECOLI	prf.2417383A
	ORF (bp)	504	924	1182	1521	366	303	627	792	786	936	213	339	663	1080	195	780	1134
	Terminal (nt)	2141760	2141763	2142885	2144066	2145576	2146264	2146566	2148022	2147261	2149166	2149359	2149634	2150997	2152118	2152329	2153113	2154191
	Initial (nt)	2141257	2142686	2144066	2145586	2145941	2146566	2147192	2147231	2148046	2148231	2149571	2149972	2150335	2151039	2152135	2152334	2153058
	SEQ NO.	5722	5723	5724	5725	5726	5727	5728	5729	5730	5731	5732	5733	5734	5735	5736	5737	5738
	SEQ NO. (DNA)	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238

	, y,	ŗ	c	0	ator													ii.				
10	Function	transcriptional accessory protein	sporulation-specific degradation regulator protein	dicarboxylase translocator	2-oxoglutarate/malate translocator	3-carboxy-cis, cis-muconate cycloisomerase				tRNA (guanine-N1)- methyltransferase	hypothetical protein	16S rRNA processing protein	hypothetical protein	30S ribosomal protein S16	sin	ABC transporter	ABC transporter	signal recognition particle protein				cell division protein
		trans	spor regu	dicar	2-ox	3-ca cyclc				tRN/ meth	hypo	16S	hypc	308	inversin	ABC	ABC	sign				le3
15	Matched length (a.a.)	9//	334	456	65	350				273	210	172	69	83	196	256	318	559				505
20	Similarity (%)	78.7	65.3	78.3	80.0	66.3				64.8	57.6	72.1	66.7	79.5	61.7	69.1	63.8	78.2				66.1
	Identity (%)	56.6	27.0	45.8	40.0	39.1				34.8	30.5	52.3	29.0	47.0	32.1	26.6	35.5	58.7				37.0
Table 1 (continued)	us gene	sis TOHAMA I	58 degA	neumoniae	chloroplast	tida pcaB				(12 trmD	elicolor A3(2)	prae M	ri J99 jhp0839	38 rpsP	,	alactiae cylB	oshii OT3 mtrA	38 ffh				.12 ftsY
Table 1	Hamologous gene	Bordetella pertussis TOHAMA I tex	Bacillus subtilis 168 degA	Chlamydophila pneumoniae CWL029 ybhl	Spinacia oleracea chloroplast	Pseudomonas putida pcaB				Escherichia coli K12 trmD	Streptomyces coelicolor A3(2) SCF81.27	Mycobacterium leprae MLCB250.34. rimM	Helicobacter pylori J99 jhp0839	Bacillus subtilis 168 rpsP	Mus musculus inv	Streptococcus agalactiae cylB	Pyrococcus horikoshii OT3 mtrA	Bacillus subtilis 168 ffh				Escherichia coli K12 ftsY
<i>35</i>	db Match	sp.TEX_BORPE	pir.A36940	pir.H72105	prf.2108268A	sp:PCAB_PSEPU	•			sp:TRMD_ECOL!	gp:SCF81_27	sp:R/MM_MYCLE	pir.B71881	pir:C47154	pir.T14151	prf:2512328G	prf:2220349C	sp:SR54_BACSU				sp:FTSY_ECOLI
	ORF (bp)	2274 s	975 p	1428 p	219 p	1251 s	66	393	069	819 s	648 g	513 s	348 p	495 p	576 p	867 p	876 p	1641 s	633	417	699	1530 s
45	Terminal (nt)	2154460	2156747	2157754	2159019	2159287	2160768	2161111	2161507	2162196	2163745	2163748	2164737	2164815	2166098	2166124	2166990	2167944	2171058	2172131	2172877	2173759
50	Initial (nt)	2156733	2157721	2159181	2159237	2160537	2160670	2161503	2162196	2163014	2163098	2164260	2164390	2165309	2165523	2166990	2167865	2169584	2170426	2171715	2172209	2175288
	SEQ NO. (a.a.)	5739	5740	5741	5742	5743	5744	5745	5746	5747	5748	5749	5750	5751	25/2	5753	5754	5755	5756	2525	5758	5759
55	SEQ NO.	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259

hypothetical protein

62.6

35.3

Streptomyces coelicolor A3(2) SC9C7.02

transport protein ABC transporter

559 541 388

55.6 58.8

26.6 28.3

Escherichia coli K12 cydC Streptomyces verticillus

1530 sp.CYDC\_ECOLI 1644 prf:2104260G

2193165 2194694

2276 5776 5777

2277

gp:SC9C7\_2

1122

2198004

2196883 2193165 2191522

5778

2278 2279

441

5779 2198447 2198007

5 10	Function			glucan 1,4-alpha-glucosidase or glucoamylase S1/S2 precursor		chromosome segregation protein	acytphosphatase		transcriptional regulator	hypothetical membrane protein			cation efflux system protein	formamidopyrimidine-DNA glycosylase	ribonuclease III	hypothetical protein	hypothetical protein	
15	Matched length (a.a.)			1144		1206	92		305	257			188	285	221	176	238	
20	Similarity (%)			46.2		72.6	73.9		0.09	73.5			9.92	66.7	76.5	62.5	76.9	
	Identity (%)			22.4		48.3	51.1		23.9	39.3			46.8	36.1	40.3	35.8	50.0	
25 (par	a			ae ae		sis	sis		~				dəl	M or		sis	sis	
% & & & & & & & & & & & & & & & & & & &	Homologous gene			Saccharomyces cerevisiae S288C YIR019C sta1		Mycobacterium tuberculosis H37Rv Rv2922c smc	Mycobacterium tuberculosis H37Rv RV2922.1C		Escherichia coli K12 yfeR	Mycobacterium leprae MLCL581.28c			Dichelobacter nodosus gep	Escherichia coli K12 mutM or fpg	Bacillus subtilis 168 rncS	Mycobacterium tuberculosis H37Rv Rv2926c	Mycobacterium tuberculosis H37Rv Rv2927c	
40	db Match			sp.AMYH_YEAST		sp:Y068_MYCTU	sp:ACYP_MYCTU		sp:YFER_ECOLI	pir:S72748			gp:DNINTREG_3	sp:FPG_ECOLI	pir.B69693	sp:Y06F_MYCTU	sp:Y06G_MYCTU	
	ORF (bp)	159	702	3393	963	3465	282	1854	858	831	183	447	615	858	741	534	789	
45	Terminal (nt)	2175888	2177103	2176110	2181880	2179628	2183110	2183405	2185351	2187129	2187342	2187233	2187692	2188313	2189166	2189906	2190540	
50	Initial (nt)	2176046	2176402	2179502	2180918	2183092	2183391	2185258	2186208	2186299	2187160	2187679	2188306	2189170	2189906	2190439	2191328	
	SEQ NO.	5760	5761	5762	5763	5764	5765	5766	5767	5768	5769	5770	5771	5772	5773	5774	5775	Ī
55	SEQ NO. (DNA)	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	

_													т т		r		
5	_			tein			rylase / ase		lyceryl	sphate te synthase	ne protein	cyclohydrolase		ate	nino-5- oxamide	sferase	stance protein ansport protein
10	Function	hypothetical protein	peptidase	sucrose transport protein			maltodextrin phosphorylase / glycogen phosphorylase	hypothetical protein	prolipoprotein diacylglyceryl transferase	indole-3-glycerol-phosphate synthase / anthranilate synthase component II	hypothetical membrane protein	phosphoribosyl-AMP cyclohydrolase	cyclase	inositol monophosphate phosphatase	phosphoribosylformimino-5- aminoimidazole carboxamide ribotide isomerase	glutamine amidotransferase	chloramphenicol resistance protein or transmembrane transport protein
15	Matched length (a.a.)	405	353	133			814	295	264	169	228	83	258	241	245	210	402
20	Similarity (%)	43.7	64.3	51.9			67.4	66.4	65.5	62.1	58.8	79.8	97.7	94.0	97.6	92.4	54.0
	Identity (%)	21.0	32.9	27.1			36.1	33.9	31.4	29.6	29.4	52.8	97.3	94.0	95.9	86.7	25.6
<i>25</i>		38	ပ	_			<u>a</u>		JA 485		sis	АТСС	E n	Wn;	um	E N	cmIR
© Table 1 (continued)	Homologous gene	Thermotoga maritima MSB8 TM0896	Campylobacter jejuni ATCC 43431 hipO	Arabidopsis thaliana SUC1			Thermococcus litoralis malP	Bacillus subtilis 168 yfiE	Staphylococcus aureus FDA 485 Igt	Emericella nidulans trpC	Mycobacterium tuberculosis H37Rv Rv1610	Rhodobacter sphaeroides ATCC 17023 hisl	Corynebacterium glutamicum AS019 hisF	Corynebacterium glutamicum AS019 impA	Corynebacterium glutamicum AS019 hisA	Corynebacterium glutamicum AS019 hisH	Streptomyces lividans 66 cmIR
35		Thermot TM0896	Camp 4343	Arabi			Therr	Bacill	Stap	Emer	Myco H37F	Rhod 1702	Cory AS01	Cory AS01	Cory AS01	Cory ASO	Strep
40	db Match	pir.A72322	sp:HIPO_CAMJE	pir:S38197			prf.2513410A	Sp: YFIE_BACSU	sp:LGT_STAAU	sp:TRPG_EMENI	pir.H70556	sp:HIS3_RHOSH	sp:HIS6_CORG	prf.2419176B	gp:AF051846_1	gp:AF060558_1	sp:CMLR_STRLI
	ORF (bp)	1284	1263	336	135	276	2550	900	948	801	657	354	774	825	738	633	1266
45	Terminal (nt)	2199758	2201070	2201073	2201450	2201594	2201992	2204591	2207302	2208367	2209232	2209920	2210273	2211051	2211882	2212641	2214321
50	Initial (nt)	2198475	2199808	2201408	2201584	2201869	2204541	2205490	2208249	2209167	2209888	2210273	2211046	2211875	2212619	2213273	2215586
	SEQ NO.	5780	5781	5782	5783	5784	5785	5786	5787	5788	5789	5790	5791	5792	5793	5794	5795
55	SEQ NO.	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295

		_																		
5			sphate		se	otein			phosphatase	:	enzyme	*		genase	ron repressor	TP-binding				
10	Function		imidazoleglycerol-phosphate dehydratase	histidinol-phosphate aminotransferase	histidinol dehydrogenase	serine-rich secreted protein			histidine secretory acid phosphatase	tet repressor protein	glycogen debranching enzyme	hypothetical protein	oxidoreductase	myo-inositol 2-dehydrogenase	galactitol utilization operon repressor	ferrichrome transport ATP-binding protein or ferrichrome ABC transporter	hemin permease	iron-binding protein	iron-binding protein	hypothetical protein
15	Matched length (a.a.)		198	362	439	342			211	204	722	258	268	343	329	246	332	103	182	113
20	Similarity (%)		81.8	29.3	85.7	54.4			2.65	8.09	75.5	76.0	55.2	6.09	64.4	68.3	71.1	0.89	9.79	73.5
	Identity (%)		52.5	57.2	63.8	27.2			29.4	28.9	47.4	20.0	29.9	35.0	30.4	32.9	36.8	30.1	34.6	38.1
<i>25</i>			3(2)	3(2)	•	nbe			P-1	P1	treX	ŝ	3(2)	,						
© Table 1 (continued)	Homologous gene		Streptomyces coelicolor A3(2) hisB	Streptomyces coelicolor A3(2) hisC	Mycobacterium smegmatis ATCC 607 hisD	Schizosaccharomyces pombe SPBC215.13			Leishmania donovani SAcP-1	Escherichia coli plasmid RP1 tetR	Sulfolobus acidocaldarius treX	Mycobacterium tuberculosis H37Rv Rv2622	Streptomyces coelicolor A3(2) SC2G5.27c gip	Sinorhizobium meliloti idhA	Escherichia coli K12 galR	Bacillus subfilis 168 fhuC	Vibrio cholerae hutC	Bacillus subtilis 168 yvrC	Bacillus subtilis 168 yvrC	Escherichia coli K12 ytfH
35			Streg	Stre hisC		Sch SPE			Leis	Esch tetR	Sulf	Myc H37	Stre SC2	Sinc	Esc		Vibr	Bac	Вас	Escl
40	db Match		sp:HIS7_STRCO	sp:HIS8_STRCO	sp:HISX_MYCSM	gp:SPBC215_13			prf:2321269A	pir.RPECR1	prf.2307203B	pir.E70572	gp:SC2G5_27	prf.2503399A	sp:GALR_ECOLI	sp:FHUC_BACSU	prf:2423441E	pir:G70046	pir:G70046	sp:YTFH_ECOLI
	ORF (bp)	225	909	1098	1326	1200	651	309	642	561	2508	801	774	1011	966	798	1038	348	594	441
45	Terminal (nt)	2215639	2215869	2216494	2217600	2220358	2220459	2221919	2221187	2222518	2225035	2225949	2225990	2226769	2228901	2229099	2229900	2230947	2231339	2232016
50	Initial (nt)	2215863	2216474	2217591	2218925	2219159	2221109	2221611	2221828	2221958	222258	2225149	2226763	2227779	2227906	2229896	2230937	2231294	2231932	2232456
	SEQ NO. (a.a.)	5796	5797	5798	5799	5800	5801	5802	5803	5804	5805	5806	5807	5808	5809	5810	5811	5812	5813	5814
55	SEQ NO.	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314

5		Function	DNA polymerase III epsilon chain		maltooligosyl trehalose synthase	hypothetical protein					alkanal monooxygenase alpha chain	hypothetical protein		maltooligosyltrehalose trehalohydrolase	hypothetical protein	threonine dehydratase			Corynebacterium glutamicum AS019	DNA polymerase III	chloramphenicol sensitive protein	histidine-binding protein precursor	hypothetical membrane protein
15		Matched length (a.a.)	355 D		814 п	322 h					375 a	120 h		568 tr	214 h	436 11			415 C	1183 D	279 cl	149 hi	198 h
20		Similarity (%)	50.1		68.6	52.8					54.4	79.2		72.4	72.4	99.3			49.6	80.5	73.8	25.7	64.7
		Identity (%)	23.4		42.0	27.6					20.5	58.3		46.3	36.5	99.3			22.7	53.3	37.6	21.5	22.7
<i>25</i>	(pai		43(2)								SL	13(2)				cum		•	Е	(2)		√2 hisJ	F2388
30	lable I (confinued)	Homologous gene	Streptomyces coelicolor A3(2) SCI8.12		Arthrobacter sp. Q36 treY	Deinococcus radiodurans DR1631					Photorhabdus luminescens ATCC 29999 luxA	Streptomyces coelicolor A3(2) SC7H2.05		Arthrobacter sp. Q36 treZ	Bacillus subtilis 168	Corynebacterium glutamicum ATCC 13032 ilvA			Catharanthus roseus metE	Streptomyces coelicolor A3(2) dnaE	Escherichia coli K12 rarD	Campylobacter jejuni DZ72 hisJ	Archaeoglobus fulgidus AF2388
40		db Match	gp:SCI8_12		pir:S65769	gp:AE002006_4					sp:LXA1_PHOLU	gp:SC7H2_5		pir:S65770	sp:YVYE_BACSU	sp:THD1_CORGL			pir.S57636	prf.2508371A	sp:RARD_ECOLI	sp.HISJ_CAMJE	pir.D69548
		ORF (bp)	1143	909	2433	1023	399	198	189	1056	1044	378	231	1785	651	1308	507	156	1203	3582	840	468	918
45		Terminal (nt)	2234070	2234763	2237284	2238353	2238694	2239845	2240058	2239508	2241724	2241738	2242129	2244819	2242393	2244864	2246892	2246295	2247006	2248358	2252856	2253659	2254642
50		Initial (nt)	2232928	2234158	2234852	2237331	2239092	2240042	2240246	2240563	2240681	2242115	2242359	2243035	2243043	2246171	2246386	2246450	2248208	2251939	2252017	2253192	2253725
		SEQ NO.	5815	5816	5817	5818	5819	5820	5821	5822	5823	5824	5825	5826	5827	5828	5829	5830	5831	5832	5833	5834	5835
55		SEQ NO.	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335

ਉ
ntinue
1 (00
Table

										_	_								
Function	short chain dehydrogenase or general stress protein	diaminopimelate (DAP) decarboxylase	cysteine synthase		ribosomal large subunit pseudouridine synthase D	lipoprotein signal peptidase		oleandomycin resistance protein		hypothetical protein	L-asparaginase	DNA-damage-inducible protein P	hypothetical membrane protein	transcriptional regulator		hypothetical protein	isoleucyl-tRNA synthetase		
Matched length (a.a.)	280	445	314		326	154		550		158	321	371	286	334		212	1066		
Similarity (%)	80.0	47.6	64.3		61.0	61.7		64.0		9.73	62.0	2.09	61.5	73.1		67.0	65.4		
Identity (%)	48.2	22.9	32.8		36.5	33.8		36.4		36.7	31.2	31.8	31.5	44.3		42.0	38.5		
Homologous gene	Bacillus subtilis 168 ydaD	Pseudomonas aeruginosa lysA	Alcaligenes eutrophus CH34 cysM		Escherichia coli K12 rluD	Pseudomonas fluorescens NCIB 10586 IspA		Streptomyces antibioticus oleB		Rhodococcus erythropolis orf17	Bacillus licheniformis	Escherichia coli K12 dinP	Escherichia coli K12 ybiF	Streptomyces coelicolor A3(2) SCF51.06		Streptomyces coelicolor A3(2) SCF51.05	Saccharomyces cerevisiae A364A YBL076C ILS1		
db Match	sp:GS39_BACSU	sp:DCDA_PSEAE	sp:CYSM_ALCEU		sp:RLUD_ECOLI	sp:LSPA_PSEFL		pir:S67863		prf:2422382P	sp:ASPG_BACLI	sp:DINP_ECOL!	sp:YBIF_ECOLI	gp:SCF51_6		gp:SCF51_5	sp:SYIC_YEAST		
ORF (bp)	978	1287	951	579	930	534	1002	1650	303	900	975	1401	828	1002	132	627	3162	216	1095
Terminal (nt)	2254683	2255738	2258362	2259421	2260002	2260934	2262689	2264499	2265298	2264509	2266394	2266897	2268388	2269260	2270435	2270258	2270988	2274473	2274767
Initial (nt)	2255558	2257024	2259312	2259999	2260931	2261467	2261688	2262850	2264996	2265108	2265420	2268297	2269245	2270261	2270304	2270884	2274149	2274688	2275861
SEO NO. (a.a.)	5836	5837	5838	5839	5840	5841	5842	5843	5844	5845	5846	5847	5848	5849	5850	5851	5852	5853	5854
SEQ NO. (DNA)	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

hypothetical protein (putative YAK 1 protein) pyrophosphoryl-undecaprenol N-acetylglucosamine pyrophosphoryl-undecaprenol N-acetylglucosamine cell division initiation protein or cell division protein UDP-N-acetylmuramoylalanine-D-glutamate ligase UDP-N-acetylmuramate--alanine ligase UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate-D-alanyl-D-alanyl ligase hypothetical membrane protein acety/muramyl-(pentapeptide) UDP-N-acetylglucosamine-Nphospho-n-acetylmuramoyl-pentapeptide Function hypothetical protein hypothetical protein cell division protein hypothetical protein cell division protein Matched length (a.a) 372 110 246 117 442 486 490 365 152 222 494 221 82 Similarity 100.0 96.6 100.0 51.0 8.66 99.5 93.6 63.8 64.2 73.2 99.3 99.1 98 Identity (%) 39.0 98.6 93.6 38.6 35.0 99.2 6 က က 97.7 9 6 99 98 99 99 Brevibacterium lactofermentum yfih Brevibacterium lactofermentum ATCC 13869 ftsW Brevibacterium lactofermentum ftsZ Brevibacterium lactofermentum ATCC 13869 murG Brevibacterium lactofermentum ATCC 13869 murD Brevibacterium lactofermentum orf6 Corynebacterium glutamicum Corynebacterium glutamicum ftsQ Corynebacterium glutamicum murC Fable 1 (continued) Mycobacterium tuberculosis Escherichia coli K12 mraY Escherichia coli K12 murF Homologous gene Mus musculus P4(21)n H37Rv Rv2146c Sp:MURF\_ECOLI gp:BLA242646\_3 Sp:MRAY\_ECOLI gp:BLA242646\_2 sp:YFZ1\_CORGL sp:FTSZ\_BRELA GP:AB028868\_1 gp:BLA242646\_1 gp:AB015023\_1 φ, db Match prf:2420425C gsp:W70502 gp:BLFTSZ\_ pir:F70578 1116 1458 1098 1542 1326 1650 ORF (bp) 285 456 663 738 486 999 468 384 333 2287969 2282661 2286655 2277416 2278122 2279640 2278890 2280470 2281166 2283782 2285437 2286862 Terminal 2276353 2276881 2286831 <u>E</u> 2283776 2289510 2280215 2281135 2282623 2285431 2278078 2285904 2286272 2286499 2287959 2277336 2276637 2278859 2279155 Œ 5865 5866 5869 5863 5864 5867 5868 5855 5856 5859 5860 5862 5857 5858 5861 SEO. (a.a.) 2365 2366 2363 2364 2367 2369 (DNA) 2355 2356 2358 2359 2360 2362 2357 2361 ò

Function  UDP-N-acetylmuramoylalanyl-D- alanyl-D-alanyl ligase penicillin-binding protein hypothetical protein hypothetical protein hypothetical protein hypothetical protein hypothetical protein hypothetical protein hypothetical protein hypothetical protein hypothetical protein bypothetical protein hypothetical protein bypothetical protein  5,10-methylenetetrahydrofolate reductase dimethylallyltranstransferase hypothetical membrane protein hypothetical protein	otein
Function UDP-N-acetylmuramoy glutamyl-2,6-diaminopia alanyl-D-alanyl ligase penicillin binding protei penicillin-binding protei hypothetical protein hypothetical protein hypothetical protein hypothetical membrane dimethylallyltranstransf hypothetical membrane dimethylallyltranstransf eukaryotic-type protain	p bid
<del></del>	hypothetical membrane protein
Matched length (a.a.) 491 491 143 137 137 157 650 650 650 6684 684 684	411
Similarity (%) (%) (%) (%) (%) (%) (%) (%) (%) (%)	58.4
100.0 100.0	30.7
nued) nued) nued) asa pbpB osa pbpB	
Table 1 (continued)  Homologous gene Homologous gene Bacillus subtilis 168 murE Brevibacterium lactofermentum ORF2 pbp Pseudomonas aeruginosa pbpB Mycobacterium leprae MLCB268.11c Mycobacterium leprae MLCB268.11c Mycobacterium leprae MLCB268.13 Streptomyces lividans 1326 metF Mycobacterium leprae MLCB268.13 Streptomyces lividans 1326 metF Mycobacterium leprae MLCB268.17 Mycobacterium leprae MLCB268.17 Mycobacterium leprae Mycobacterium leprae Mycobacterium leprae Mycobacterium leprae Mycobacterium leprae Mycobacterium leprae Mycobacterium tuberculosis H37Rv Rv2175c Streptomyces coelicolor A3(2) pkaF	Mycobacterium leprae MLCB268.23
db Match  sp:MURE_BACSU  GSP:Y33117  pir.S54872  pir.A70581  sp:MLCB268_11  sp:MLCB268_13  sp:MLCB268_16  pir.S32168  gp:MLCB268_16  pir.A70936  pir.A70936	gp:MLCB268_21
ORF (bp) 1551 1551 1953 795 1011 429 387 429 978 978 573 567 567 567 567 567 567 567 567 567 567	651
Terminal (nt) (nt) (nt) (nt) (nt) (nt) 2289523 2290973 2294117 2296512 2299431 2298438 2298451 2302685 2302685 2304980 2304980	2303040 2306218
1011 (nt) (nt) (2291073 2291197 2293164 2294117 2295804 2297653 22996898 22996898 22996898 22996898 22996898 22996898 22996898 2300706 2300706 2302833	2303690
SEQ NO. (a.a.) 5870 5871 5875 5875 5875 5878 5878 5880 5880 5888	5885 5886
SEQ NO. (DNA) 2370 2376 2378 2380 2381 2383 2384 2384 2388	2385

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

-																
	Function	hypothetical membrane protein	3-deoxy-D-arabino-heptulosonate-7-phosphate synthase	hypothetical protein	hypothetical membrane protein	major secreted protein PS1 protein precursor			hypothetical membrane protein	acyltransferase	glycosyl transferase	protein P60 precursor (invasion- associated-protein)	protein P60 precursor (invasion- associated-protein)	ubiquinol-cytochrome c reductase cytochrome b subunit	ubiquinol-cytochrome c reductase iron-sulfur subunit (Rieske [eFe-2S] iron-sulfur protein cyoB	ubiquinol-cytochrome c reductase cytochrome c
	Matched length (a.a.)	434	462	166	428	440			249	245	383	296	191	201	203	278
	Similarity (%)	62.0	6.78	77.7	64.5	57.1			100.0	100.0	75.7	8.09	61.3	64.7	57.1	83.1
	Identity (%)	30.4	6.99	58.4	35.1	28.2			100.0	100.0	50.1	26.4	33.0	34.3	37.9	58.6
Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv2181	Amycolatopsis mediterranei	Mycobacterium leprae MLCB268.21c	Mycobacterium tubercutosis H37Rv Rv2181	Corynebacterium glutamicum (Brevibacterium flavum) ATCC 17965 csp1			Corynebacterium glutamicum ATCC 13032	Corynebacterium glutamicum ATCC 13032	Streptomyces coelicolor A3(2) SC6G10.05c	Listeria ivanovii iap	Listeria grayi iap	Heliobacillus mobilis petB	Streptomyces lividans qcrA	Mycobacterium tuberculosis H37Rv Rv2194 qcrC
	db Match	pir.G70936	gp:AF260581_2	gp:MLCB268_20	pir:G70936	sp.CSP1_CORGL			gp:AF096280_3	gp:AF096280_2	gp:SC6G10_5	sp:P60_LISIV	sp:P60_LISGR	prf.2503462K	gp:AF107888_1	sp:Y005_MYCTU
	ORF (bp)	1308	1386	504	2418	1449	204	177	1188	735	1143	1047	627	1602	672	885
	Terminal (nt)	2307621	2307697	2309173	2312252	2313808	2314036	2313916	2314236	2315678	2317633	2318804	2319968	2321472	2323088	2324311
	Initial (nt)	2306314	2309082	2309676	2309835	2312360	2313833	2314092	2315423	2316412	2318775	2319850	2320594	2323073	2323759	2325195
	SEQ NO. (a.a.)	5887	5888	5889	5890	5891	5892	5893	5894	5895	5896	5897	5898	5899	5900	5901
	SEQ NO. (DNA)	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401

5	
10	
15	
20	•
25	
30	
35	
40	
45	
50	

Table 1 (continued)

																_		
Function	cytochrome c oxidase subunit III		hypothetical membrane protein	cytochrome c oxidase subunit II	glutarnine-dependent amidotransferase or asparagine synthetase (lysozyme insensitivity protein)	hypothetical protein	hypothetical membrane protein	cobinamide kinase	nicotinate-nucleotide— dimethylbenzimidazole phosphoribosyltransferase	cobalamin (5'-phosphate) synthase		clavulanate-9-aldehyde reductase	branched-chain amino acid aminotransferase	leucyl aminopeptidase	hypothetical protein	dihydrolipoamide acetyltransferase		lipoyltransferase
Matched length (a.a.)	188		145	317	640	114	246	172	341	305		241	364	493	97	691		210
Identity Similarity (%)	70.7		71.0	53.9	99.8	100.0	60.2	64.0	6.99	49.8		68.5	70.3	62.9	0.79	68.5		65.7
Identity (%)	36.7		38.6	28.7.	99.7	100.0	35.0	43.0	37.8	25.3		38.6	40.1	36.3	40.2	48.9		36.7
Homologous gene	Synechococcus vulcanus		Mycobacterium tuberculosis H37Rv Rv2199c	Rhodobacter sphaeroides ctaC	Corynebacterium glutamicum KY9611 ItsA	Corynebacterium glutamicum KY9611 orf1	Mycobacterium leprae MLCB22.07	Rhodobacter capsulatus cobP	Pseudomonas denitrificans cobU	Pseudomonas denitrificans cobV		Streptomyces clavuligerus car	Mus musculus BCAT1	Pseudomonas putida ATCC 12633 pepA	Saccharopolyspora erythraea ORF1	Streptomyces seoulensis pdhB		Arabidopsis thaliana
db Match	Sp.COX3_SYNVU		sp.Y00A_MYCTU	sp:COX2_RHOSH	gp:AB029550_1	gp:AB029550_2	gp:MLCB22_2	pir.S52220	sp:COBU_PSEDE	sp:COBV_PSEDE		prf:2414335A	sp:ILVE_MYCTU	gp:PPU010261_1	prf:2110282A	gp:AF047034_2		gp:AB020975_1
ORF (bp)	615	153	429	1077	1920	342	768	522	1089	921	237	714	1137	1500	393	2025	1365	753
Terminal (nt)	2325273	2326121	2326472	2326921	2330435	2330586	2331967	2332495	2333600	2334535	2334481	2335028	2335915	2338734	2338748	2341293	2339440	2342164
Initial (nt).	2325887	2326273	2326900	2327997	2328516	2330927	2331200	2331974	2332512	2333615	2334717	2335741	2337051	2337235	2339140	2339269	2340804	2341412
SEQ NO. (a.a.)	5902	5903	5904	5905	5906	5907	5908	6069	5910	5911	5912	5913	5914	5915	5916	5917	5918	5919
SEQ NO. (DNA)	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419

4-hydroxyphenylacetate permease transmembrane transport protein

433

53.4 72.8

21.9 42.4

Escherichia coli hpaX

2434

158

Streptomyces coelicolor A3(2) SCGD3.10c

gp:SCGD3\_10

561

2357354

2356794

5935

transmembrane transport protein

118

66.1

31.4

Streptomyces coelicolor A3(2) SCGD3.10c

9p:SCGD3\_10

444

2357707

2436 2437

195 405

2357290 2358130

2438

										,							_
5		E		ne protein	ne protein			ne protein		protein			ise alpha chain Ipha chain)	oitor nhibitor)			
10		Function	lipoic acid synthetase	hypothetical membrane protein	hypothetical membrane protein	transposase (ISCg2)		hypothetical membrane protein		mutator mutT domain protein	hypothetical protein		alkanal monooxygenase alpha chain (bacterial luciferase alpha chain)	protein synthesis inhibitor (translation initiation)			
15		Matched length (a.a.)	285	257	559	401		157		145	128		220	111			
20		Similarity (%)	70.9	76.7	8.79	100.0		63.7		44.0	65.6		6.09	73.0			
		Identity (%)	44.6	45.5	32.9	100.0	!	41.4		31.0	36.7		25.0	40.5			· ,
25	Table 1 (continued)	Homologous gene	nolicus GRA BD	uberculosis	K12 yidE	ı glutamicum		elicolor A3(2)			itima MSB8		♦	itima MSB8			
35	Table 1	Нотою	Pelobacter carbinolicus GRA BD 1 lipA	Mycobacterium tuberculosis H37Rv Rv2219	Escherichia coli K12 yidE	Corynebacterium glutamicum ATCC 13032 tnp		Streptomyces coelicolor A3(2) SC5F7.04c			Thermotoga maritima MSB8 TM1010		Vibrio harveyi luxA	Thermotoga maritima MSB8 TM0215			
40		db Match	p:LIPA_PELCA	p:Y00U_MYCTU	p.YIDE_ECOLI	p:AF189147_1		gp:SC5F7_34			pir:B72308		p:LUXA_VIBHA	pir.A72404			
		ORF (bp)	1044 sp	ds 087	1617 sp	1203 gp	300	471 gp	213	975	399 pir	009	849 sp	393 pir	243	261	-
<b>45</b> .		Terminal (nt)	2343347	2344258	2346047	2346289	2347804	2348078	2350408	2351996	2350912	2351310	2352828	2353225	2355398	2355180	
50		Initial (nt)	2342304	2343479	2344431	2347491	2347505	2348548	2350620	2351022	2351310	2351909	2351980	2352833	2355156	2355440	
		SEQ NO. (a.a.)	5920	5921	5922	5923	5924	5925	5926	5927	5928	5929	5930	5931	5932	5933	Ĺ
55		SEQ NO. (DNA)	2420	2421	2422	2423	2424	2425	2426	2427	2428	2429	2430	2431	2432	2433	
																	-

5		
10		
15		
20		
25		
30		
35	•	
40		
45		
50		

	Function		heme oxygenase	glutamate-ammonia-ligase adenylyltransferase	glutamine synthetase	hypothetical protein	hypothetical protein	hypothetical protein	galactokinase	virulence-associated protein		bifunctional protein (ribonuclease H and phosphoglycerate mutase)		hypothetical protein	hypothetical protein	phosphoglycolate phosphatase	low molecular weight protein- tyrosine-phosphatase	hypothetical protein	insertion element (IS402)
	Matched length (a.a.)		214	809	441	392	601	54	374	358		382		249	378	204	156	281	129
	Similarity (%)		78.0	67.0	73.0	54.1	58.2	55.6	53.7	54.5		75.1		58.6	76.2	54.4	63.5	65.5	56.6
	Identity (%)		57.9	43.4	43.5	26.8	33.4	38.9	24.9	27.1		54.7		26.5	49.2	26.0	46.2	40.9	32.6
Table 1 (continued)	Homologous gene		Corynebacterium diphtheriae C7 hmuO	Streptomyces coelicolor A3(2) ginE	Thermotoga maritima MSB8 glnA	Streptomyces coelicolor A3(2) SCE9.39c	Mycobacterium tuberculosis H37Rv Rv2226	Streptomyces coelicolor A3(2) SCC75A.11c.	Homo sapiens galK1	Brucella abortus vacB		Mycobacterium tuberculosis H37Rv Rv2228c		Mycobacterium tuberculosis H37Rv Rv2229c	Mycobacterium tuberculosis H37Rv Rv2230c	Escherichia coli K12 gph	Streptomyces coelicolor A3(2) SCQ11.04c ptpA	Mycobacterium tuberculosis H37Rv Rv2235	Burkholderia cepacia
	db Match		sp:HMUO_CORDI	gp:SCY17736_4	sp:GLNA_THEMA	gp:SCE9_39	sp:Y017_MYCTU	gp:SCC75A_11	sp:GAL1_HUMAN	gp:AF174645_1		sp:Y019_MYCTU		sp:Y01A_MYCTU	sp:Y01B_MYCTU	sp:GPH_ECOLI	sp:PTPA_STRCO	sp:Y01G_MYCTU	sp.YI21_BURCE
	ORF (bp)	543	645	3135	1338	1104	1827	180	1293	1266	486	1146	729	717	1140	654	471	954	393
	Terminal (nt)	2358153	2358772	2359614	2362818	2365455	2367413	2367473	2369083	2369116	2370908	2371412	2373289	2372573	2373323	2375197	2375684	2376720	2376998
	Initial (nt)	2358695	2359416	2362748	2364155	2364352	2365587	2367652	2367791	2370381	2370423	2372557	2372561	2373289	2374462	2374544	2375214	2375767	5956 2377390
	SEQ NO. (a.a.)	5939	5940	5941	5942	5943	5944	5945	5946	5947	5948	5949	5950	5951	5952	5953	5954	5955	
	SEQ NO. (DNA)	2439	2440	2441	2442	2443	2444	2445	2446	2447	2448	2449	2450	2451	2452	2453	2454	2455	2456

5	٠	
10		
15		
20		
25		ed)
30		Table 1 (continued)
35		គ្ន
40		
45	·	
50		

_														—,					$\overline{}$
	Function		transcriptional regulator		hypothetical protein		pyruvate dehydrogenase component		ABC transporter or glutamine transport ATP-binding protein		ribose transport system permease protein	hypothetical protein	calcium binding protein		lipase or hydrolase	acyl carier protein	N-acetylglucosamine-6-phosphate deacetylase	hypothetical protein	
	Matched length (a.a.)		135		134		910		261		283	286	125		352	75	253	289	
	Similarity (%)		57.8		9.77		78.9		62.8		58.7	62.9	55.2		55.7	0.08	75.5	65.7	
	Identity (%)		30.4		55.2		55.9		33.7		25.4	26.2	41.6		29.6	42.7	43.9	33.6	
(2021)	Homologous gene		Streptomyces caelicolor A3(2) SC8F4.22c		Mycobacterium tuberculasis H37Rv Rv2239c		Streptomyces seoulensis pdhA		Escherichia coli K12 glnQ		Bacillus subtilis 168 rbsC	Rickettsia prowazekii Madrid E RP367	Dictyostelium discoideum AX2 cbpA		Streptomyces coelicolor A3(2) SC6G4.24	Myxococcus xanthus ATCC 25232 acpP	Escherichia coli K12 nagD	Deinococcus radiodurans DR1192	
	db Match		gp:SC8F4_22		sp:Y01K_MYCTU		gp:AF047034_4		sp.GLNQ_ECOLI		sp:RBSC_BACSU	pir.H71693	sp.CBPA_DICDI		gp:SC6G4_24	sp:ACP_MYXXA	sp:NAGD_ECOLI	gp:AE001968_4	
	ORF (bp)	243	378	198	429	345	2712	1476	789	963	888	939	810	372	1014	291	825	1032	471
	Terminal (nt)	2377484	2378276	2378489	2378884	2379770	2382744	2380765	2382827	2385426	2383622	2384509	2386580	2385913	2386614	2387957	2388821	2389869	2390434
	Initial (nt)	2377726	2377899	2378292	2379312	2379426	2380033	2382240	2383615	2384464	2384509	2385447	2385771	2386284	2387627	2387667	2387997	2388838	2390904
	SEQ NO. (a.a.)	5957	5958	5959	5960	5961	5965	5963	5964	5965	9969	5967	5968	5969	5970	5971	5972	5973	5974
	SEQ NO. (DNA)	2457	2458	2459	2460	2461	2462	2463	2464	2465	2466	2467	2468	2469	2470	2471	2472	2473	2474

5	Function	hypothetical protein						alkaline phosphatase D precursor		hypothetical protein	hypothetical protein		DNA primase	ribonuclease Sa			L-glutamine: D-fructose-6-phosphate amidotransferase			deoxyguanosinetriphosphate triphosphohydrolase	hypothetical protein
15	Matched length (a.a.)	271						530		594	68		633	98			636			414	171
20	Similarity (%)	75.3						64.7		73.1	72.1		82.9	67.4			82.2			76.3	2.69
	Identity (%)	52.4						34.2		44.4	41.2		59.1	49.0			59.1			54.6	30.4
25 (panunu	gene	olor A3(2)						рһоО		olor A3(2)	rculosis		gmatis	faciens BMK			gmatis			gmatis dgt	is NMA0251
Table 1 (continued)	Homologous gene	Streptomyces coelicolor A3(2) SC4A7.08						Bacillus subtilis 168 phoD		Streptomyces coelicolor A3(2) SCI51.17	Mycobacterium tuberculosis H37Rv Rv2342		Mycobacterium smegmatis dnaG	Streptomyces aureofaciens BMK			Mycobacterium smegmatis mc2155 glmS			Mycobacterium smegmatis dgt	Neisseria meningitidis NMA0251
40	db Match	gp:SC4A7_8						sp:PPBD_BACSU		gp:SCI51_17	pir:G70661		prf:2413330B	gp:XXU39467_1			gp:AF058788_1			prf:2413330A	gp:NMA1Z2491_23
	ORF (bp)	825	492	771	546	465	342	1560	714	1836	240	675	1899	462	243	636	1869	324	1152	1272	675
45	Terminal (nt)	2391184	2392075	2392579	2393970	2393973	2394935	2396763	2395273	2399099	2399397	2399668	2399405	2401834	2402080	2402530	2402144	2404846	2406822	2404987	2406262
50	Initial (nt)	2392008	2392566	2393349	2393425	2394437	2394594	2395204	2395986	2397264	2399158	2400342	2401303	2401373	2401838	2403165	2404012	2404523	2405671	2406258	2406936
·	SEQ NO. (a.a.)	5975	5976	5977	8269	5979	5980	5981	5982	5983	5984	5985	5986	5987	5988	5989	2990	5991	5992	5993	5994
55	SEQ NO.	2475	2476	2477	2478	2479	2480	2481	2482	2483	2484	2485	2486	2487	2488	2489	2490	2491	2492	2493	2494

5
10
15
20
25
30
35
40
45
50

	Function	hypothetical protein	hypothetical protein		glycyl-tRNA synthetase	bacterial regulatory protein, arsR family	ferric uptake regulation protein	hypothetical protein (conserved in C.glutamicum?)	hypothetical membrane protein	undecaprenyl diphosphate synthase	hypothetical protein	Era-like GTP-binding protein	hypothetical membrane protein	hypothetical protein	Neisserial polypeptides predicted to be useful antigens for vaccines and diagnostics	phosphate starvation inducible protein	hypothetical protein	
	Matched length (a.a.)	692	138		208	89	132	529	224	233	245	296	432	157	85	344	248	
	Similarity (%)	63.6	54.4		6.69	73.0	5.07	46.7	67.0	71.2	74.3	70.3	82.4	86.0	50.0	84.6	75.4	
	identity (%)	31.1	24.6		46.1	49.4	34.9	24.8	40.6	43.4	45.7	39.5	52.8	65.0	45.0	61.1	44.0	
Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv2345	Drosophila melanogaster CG10592		Thermus aquaticus HB8	Mycobacterium tuberculosis H37Rv Rv2358 furB	Escherichia coli K12 fur	Mycobacterium tuberculosis H37Rv Rv1128c	Streptomyces coelicolor A3(2) h3u	Micrococcus luteus 8-P 26 uppS	Mycobacterium tuberculosis H37Rv Rv2362c	Streptococcus pneumoniae era	Mycobacterium tuberculosis H37Rv Rv2366	Mycobacterium tuberculosis H37Rv Rv2367c	Neisseria meningitidis	Mycobacterium tuberculosis H37Rv Rv2368c phoH	Streptomyces coelicolor A3(2) SCC77.19c.	
	db Match	pir.B70662	gp:AE003565_26		pir:S58522	pir.E70585	sp:FUR_ECOLI	pir.A70539	gp:AF162938_1	sp:UPPS_MICLU	pir.A70586	gp:AF072811_1	sp:Y1DE_MYCTU	sp:YN67_MYCTU	GSP:Y75650	sp:PHOL_MYCTU	gp:SCC77_19	
	ORF (bp)	2037	486	582	1383	369	432	1551	792	729	726	915	1320	588	264	1050	723	942
	Terminal (nt)	2409029	2409779	2410280	2410956	2412948	2413423	2415118	2415298	2416371	2417222	2417969	2418990	2420313	2421236	2420900	2421975	2423791
	Initial (nt)	2406993	2410264	2410861	2412338	2412580	2412992	2413568	2416089	2417099	2417947	2418883	2420309	2420900	2420973	2421949	2422697	2422850
	SEQ NO. (a.a.)	5995	5996	5997	5998	5999	0009	6001	6002	6003	6004	6005	9009	6007	6008	6009	6010	6011
	SEQ NO.	2495	2496	2497	2498	2499	2500	2501	2502	2503	2504	2505	2506	2507	2508	2509	2510	2511

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

_		$\overline{}$				_									- 7			
	Function	heat shock protein dnaJ	heat-inducible transcriptional repressor (groEL repressor)	oxygen-independent coproporphyrinogen III oxidase	agglutinin attachment subunit precursor			long-chain-fatty-acidCoA ligase	4-alpha-glucanotransferase	ABC transporter, Hop-Resistance protein	Neisserial polypeptides predicted to be useful antigens for vaccines and diagnostics	polypeptides predicted to be useful antigens for vaccines and diagnostics			peptidyl-dipeptidase	carboxylesterase	glycosyl hydrolase or trehalose synthase	hypothetical protein
	Matched length (a.a.)	380	334	320	134		·	611	738	604	89	. 107			069	453	594	449
	Similarity (%)	77.4	79.6	64.1	64.9			75.1	55.4	64.4	51.0	53.0			68.3	45.7	84.9	58.8
	Identity (%)	47.1	48.2	33.1	36.6			48.0	28.3	29.5	44.0	47.0			40.3	24.1	65.2	32.1
Table 1 (continued)	Homologous gene	Streptomyces albus dnaJ2	Streptomyces albus hrcA	Bacillus stearothermophilus hemN	Saccharomyces cerevisiae YNR044W AGA1			Streptomyces coelicolor A3(2) SC6G10.04	Escherichia coli K12 malQ	Lactobacillus brevis plasmid horA	Neisseria gonorrhoeae	Neisseria meningitidis			Salmonella typhimurium dcp	Anisopteromalus calandrae	Mycobacterium tuberculosis H37Rv Rv0126	Mycobacterium tuberculosis H37Rv Rv0127
	db Match	prf:2421342B	prf.2421342A	prf.2318256A	sp:AGA1_YEAST			gp:SC6G10_4	sp:MALQ_ECOLI	gp:AB005752_1	GSP:Y74827	GSP:Y74829			sp:DCP_SALTY	gp:AF064523_1	pir.G70983	pir.H70983
	ORF (bp)	1146	1023	066	519	693	378	1845	2118	1863	255	333	180	204	2034	1179	1794	1089
	Terminal (nt)	2422700	2423915	2424965	2426699	2426776	2427807	2428184	2432413	2434370	2433614	2433875	2434440	2434573	2434805	2438049	2439906	2440994
	Initial (nt)	2423845	2424937	2425954	2426181	2427468	2428184	2430028	2430296	2432508	2433868	2434207	2434619	2434776	2436838	2436871	2438113	2439906
	SEQ NO. (a.a.)	6012	6013	6014	6015	6016	6017	6018	6019	6020	6021	6022	6023	6024	6025	6026	6027	6028
	SEQ NO. (DNA)	2512	2513	2514	2515	2516	2517	2518	2519	2520	2521	2522	2523	2524	2525	2526	2527	2528

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

Table 1 (continued)

Function	isopentenyl-diphosphate Delta- isomerase						beta C-S lyase (degradation of aminoethylcysteine)	branched-chain amino acid transport system carrier protein (isoleucine uptake)	alkanal monooxygenase alpha chain		malonate transporter	glycolate oxidase subunit	transcriptional regulator		hypothetical protein		heme-binding protein A precursor (hemin-binding lipoprotein)	oligopeptide ABC transporter (permease)	dipeptide transport system permease protein	oligopeptide transport ATP-binding protein
Matched length (a.a.)	189						325	426	343		324	483	203		467		546	315	27.1	372
Similarity (%)	57.7						100.0	100.0	49.0		60.5	55.1	65.0		57.6		55.5	73.3	74.5	66.4
Identity (%)	31.8				į		99.4	99.8	21.6		25.9	27.7	25.6		22.5		27.5	40.0	43.2	37.4
Homologous gene	Chlamydomonas reinhardtii ipi1						Corynebacterium glutamicum ATCC 13032 aecD	Corynebacterium glutamicum ATCC 13032 brnQ	Vibrio harveyi luxA		Sinorhizobium meliloti mdcF	Escherichia coli K12 glcD	Escherichia coli K12 ydfH		Salmonella typhimurium ygiK		Haemophilus influenzae Rd H10853 hbpA	Bacillus subtilis 168 appB	Escherichia coli K12 dppC	Escherichia coli K12 oppD
db Match	pir.T07979			-			gp.CORCSLYS_1	sp.BRNQ_CORGL	sp:LUXA_VIBHA		gp:AF155772_2	sp:GLCD_ECOLI	Sp:YDFH_ECOLI		sp:YGIK_SALTY		sp:HBPA_HAEIN	sp:APPB_BACSU	sp:DPPC_ECOLI	prf:2306258MR
ORF (bp)	585	222	438	1755	999	519	975	1278	978	522	927	2844	711	282	1347	423	1509	996	828	1437
Terminal (nt)	2441005	2441890	2442792	2441602	2443356	2444033	2445709	2446993	2447998	2450323	2450859	2451794	2455435	2455452	2455720	2457337	2459371	2460336	2461167	2462599
Initial (nt)	2441589	2441669	2442355	2443356	2444015	2444551	2444735	2445716	2447021	2450844	2451785	2454637	2454725	2455733	6043 2457066	2457759	2457863	2459371	2460340	2461163
SEQ NO. (a.a.)	6059	6030	6031	6032	6033	6034	6035	9609	6037	6038	6039	6040	6041	6042	6043	6044	60,45	6046	6047	6048
SEQ NO. (DNA)	2529	2530	2531	2532	2533	2534	2535	2536	2537	2538	2539	2540	2541	2542	2543	2544	2545	2546	2547	2548

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

	Function	hypothetical protein	hypothetical protein	lase	hypothetical membrane protein		sodium-dependent transporter or odium Bile acid symporter family	apospory-associated protein C	-	thiamine biosynthesis protein x	hypothetical protein	glycine betaine transporter				large integral C4-dicarboxylate membrane transport protein	small integral C4-dicarboxylate membrane transport protein	C4-dicarboxylate-binding periplasmic protein precursor		GTP-binding protein
		hypotheti	hypotheti	ribose kinase	hypotheti		sodium-d odium Bil	apospory		thiamine	hypothet	glycine b				farge inte	small inte membrar	C4-dicari periplasn	extensin	GTP-bin
	Matched length (a.a.)	106	157	300	466		284	295		133	197	601		•		448	118	227	46	603
	Similarity (%)	44.0	58.0	65.0	64.6		61.6	51.2		100.0	65.5	71.7			l	71.9	73.7	29.0	73.0	83.6
	identify (%)	35.0	29.3	41.0	39.9		31.3	28.5		100.0	42.6	39.8				34.6	33.9	28.2	63.0	58.7
Table 1 (continued)	Homologous gene	Aeropyrum pernix K1 APE1580	Aquifex agolicus VF5 aq 768	Rhizobium etli rbsK	Streptomyces coelicolor A3(2) SCM2.16c		Homo sapiens	Chlamydomonas reinhardtii		Corynebacterium glutamicum ATCC 13032 thiX	Mycobacteriophage D29 66	Corynebacterium glutamicum ATCC 13032 betP				Rhodobacter capsulatus dctM	Klebsiella pneumoniae dctQ	Rhodobacter capsulatus B10 dctP	Lycopersicon esculentum (tomato)	Bacillus subtilis 168 lepA
	db Match	PIR:G72536	pir:D70367	prf.2514301A	gp:SCM2_16		sp:NTCI_HUMAN	gp:AF195243_1		sp:THIX_CORGL	sp:VG66_BPMD	sp:BETP_CORGL				prf.2320266C	gp:AF186091_1	sp:DCTP_RHOCA	PRF:1806416A	sp:LEPA_BACSU
	ORF (bp)	507	549	903	1425	303	972	846	366	570	588	1890	966	1608	384	1311	480	747	243	1845
	Terminal (nt)	2461543	2462602	2464143	2465768	2465465	2466038	2467922	2470678	2472819	2472893	2475542	2477492	2479251	2479762	2479898	2481213	2481734	2484087	2482548
	Initial (nt)	2462049	6050 2463150	2463241	2464344	2465767	2467009	2467077	2470313	2472250	2473480	2473653	2476497	2477644	2479379	2481208	2481692	2482480	2483845	2484392
	SEQ NO.	6049		6051	6052	6053	6054	6055	9509	6057	6058	6909	0909	6061	6062	6063	6064	6065	9909	2909
	SEQ NO. (DNA)	2549	2550	2551	2552	2553	2554	2555	2556	2557	2558	2559	2560	2561	2562	2563	2564	2565	2566	2567

. 5		Function	nie	otein S20	protein	in	sin	late competence operon required for DNA binding and uptake	ate competence operon required for DNA binding and uptake	
10		Fur	hypothetical protein	30S ribosomal protein S20	thrreonine efflux protein	ankyrin-like protein	hypothetical protein	late competence operon DNA binding and uptake	late competence operon DNA binding and uptake	
15		Matched length (a.a.)	185	85	210	129	313	527	195	
20		Identity Similarity (%)	69.7	72.9	67.1	80.6	74.1	49.7	63.6	
		Identity (%)	41.6	48.2	30.0	61.2	46.0	21.4	30.8	
<b>25</b>	Table 1 (continued)	us gene	berculosis	12 rpsT	12 rhtC	ilicolor A3(2)	berculosis	38 comEC	38 comEA	
30	Table 1 (	Homologous gene	Mycobacterium tuberculosis H37Rv Rv2405	Escherichia coli K12 rpsT	Escherichia coli K12 rhtC	Streptomyces coelicolor A3(2) SC6D7.25.	Mycobacterium tuberculosis H37Rv Rv2413c	Bacillus subtilis 168 comEC	Bacillus subtilis 168 comEA	
35		atch							sp:CME1_BACSU	
40		db Match	pir.H70683	sp.RS20_ECOLI	sp:RHTC_ECOLI	gp:SC6D7_25	pir.H70684	sp.CME3_BACSU	sp:CME1	
		ORF (bp)	609	261	699	405	975	1539	582	822
45		Terminal (nt)	2485269	2485733	2485801	2486477	2486910	2487912	2489573	2491732
50		Initial (nt)	2484661	2485473	2486469	6071 2486881	2487884	2489450	2490154	6075 2490911
		SEQ NO. (a.a.)	8909	6909	0209	6071	6072	6073	6074	
		0 0 €	99	69	70	71	72	73	74	75

	Function	hypothetical protein	30S ribosomal protein S20	thrreonine efflux protein	ankyrin-like protein	hypothetical protein	late competence operon required for DNA binding and uptake	late competence operon required for ONA binding and uptake		hypothetical protein	phosphoglycerate mutase	hypothetical protein	hypothetical protein		gamma-glutamyl phosphate reductase or glutamate-5- semialdehyde dehydrogenase	D-isomer specific 2-hydroxyacid dehydrogenase		GTP-binding protein
	Matched length (a.a.)	185	85	210	129	313	527	195		273	235	117	197		432	304		487
	Similarity (%)	69.7	72.9	1.79	80.6	74.1	49.7	63.6		€.99	66.4	86.3	85.3		99.8	100.0		78.2
	Identity (%)	41.6	48.2	30.0	61.2	46.0	21.4	30.8		34.8	46.8	55.6	68.0		99.1	99.3		58.9
iable i (colimined)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv2405	Escherichia coli K12 rpsT	Escherichia coli K12 rhtC	Streptomyces coelicolor A3(2) SC6D7.25.	Mycobacterium tuberculosis H37Rv Rv2413c	Bacillus subtilis 168 comEC	Bacillus subtilis 168 comEA		Streptomyces coelicolor A3(2) SCC123.07c.	Mycobacterium tuberculosis H37Rv Rv2419c	Mycobacterium tuberculosis H37Rv Rv2420c	Streptomyces coelicolor A3(2) SCC123.17c.		Corynebacterium glutamicum ATCC 17965 proA	Corynebacterium glutamicum ATCC 17965 unkdh		Streptomyces coelicolor A3(2) obg
	db Match	pir:H70683	sp:RS20_ECOLI	sp:RHTC_ECOLI	gp:SC6D7_25	pir.H70684	sp.CME3_BACSU	sp:CME1_BACSU		gp:SCC123_7	pir:F70685	pir:G70685	gp:SCC123_17		sp:PROA_CORGL	sp:YPRA_CORGL		gp:D87915_1
	ORF (bp)	609	261	699	405	975	1539	582	822	822	708	471	678	1023	1296	912	711	1503
	Terminal (nt)	2485269	2485733	2485801	2486477	2486910	2487912	2489573	2491732	2490290	2491151	2491873	2492501	2493215	2494339	2495696	2497513	2498009
	Initial (nt)	2484661	2485473	2486469	2486881	2487884	2489450	2490154	2490911	2491111	2491858	2492343	2493178	2494237	2495634	2496607	2496803	6084 2499511
	SEQ NO. (a.a.)	6068	6909	0209	6071	6072	6073	6074	6075	9209	6077	8209	6009	6080	6081.	6082	6083	6084
	SEQ NO. (DNA)	2568	2569	2570	2571	2572	2573	2574	2575	2576	2577	2578	2579	2580	2581	2582	2583	2584

5
10
15
20
25
30
35
40
45
50

Table 1 (continued)

_																				
	Function	xanthine permease	2,5-diketo-D-gluconic acid reductase			50S ribosomal protein L27	50S ribosomal protein L21	ribonuclease E				hypothetical protein	transposase (insertion sequence IS31831)	hypothetical protein	hypothetical protein	nucleoside diphosphate kinase		hypothetical protein	hypothetical protein	hypothetical protein
	Matched length (a.a.)	422	276			81	101	988				195	436	117	143	134		92	112	118
	Similarity (%)	77.3	81.9			92.6	82.2	56.6				82.6	100.0	76.9	67.8	89.6		67.4	64.3	68.6
	Identity (%)	39.1	61.2			80.3	56.4	30.1				61.0	99.1	51.3	37.8	70.9		34.8	36.6	33.9
	Homologous gene	Bacillus subtilis 168 pbuX	Corynebacterium sp. ATCC 31090			Streptomyces griseus IFO13189 rpmA	Streptomyces griseus IFO13189 obg	Escherichia coli K12 rne				Streptomyces coelicolor A3(2) SCF76.08c	Corynebacterium glutamicum ATCC 31831	Streptomyces coelicolor A3(2) SCF76.08c	Streptomyces coelicolor A3(2) SCF76.09	Mycobacterium smegmatis ndk		Deinococcus radiodurans R1 DR1844	Mycobacterium tuberculosis H37Rv Rv1883c	Mycobacterium tuberculosis H37Rv Rv2446c
	db Match	sp.PBUX_BACSU	pir.140838			sp:RL27_STRGR	prf:2304263A	sp:RNE_ECOLI				gp:SCF76_8	pir:S43613	gp:SCF76_8	gp:SCF76_9	gp:AF069544_1		gp:AE002024_10	pir:H70515	pir.E70863
	ORF (bp)	1887	843	621	396	264	303	2268	549	573	747	609	1308	378	450	408	360	342	465	423
	Terminal (nt)	2501669	2501735	2503355	2504265	2503984	2504300	2504831	2507663	2507710	2508840	2509530	2509523	2511423	2511876	2511949	2512409	2513144	2513154	2513692
	Initial (nt)	2499783	2502577	2502735	2503870	2504247	2504602	2507098	2507115	2507138	2508094	2508922	2510830	2511046	2511427	2512356	2512768	2512803	2513618	2514114
	SEQ NO.	6085	6086	6087	6088	6009	0609	6091	6092	6093	6094	6095	9609	2609	8609	6609	6100	6101	6102	6103
	SEQ NO. DNA)	2585	2586	2587	2588	2589	2590	2591	2592	2593	2594	2595	2596	2597	2598	2599	2600	2601	2602	2603

5	
10	
· 15	
20	
25	
<b>30</b>	
35	
40	
45	
50	

Fable 1 (continued)
<b>H</b>

					_		,	,				· · ·							
Function	folyl-polyglutamate synthetase				valyl-tRNA synthetase	oligopeptide ABC transport system substrate-binding protein	heat shock protein dnaK	lysine decarboxylase	malate dehydrogenase	transcriptional regulator	hypothetical protein	vanillate demethylase (oxygenase)	pentachlorophenol 4- monooxygenase reductase	transport protein	malonate transporter	class-III heat-shock protein or ATP-dependent protease	hypothetical protein	succinyl CoA:3-oxoadipate CoA transferase beta subunit	succinyl CoA.3-oxoadipate CoA transferase alpha subunit
Matched length (a.a.)	451				915	521	508	170	319	207	208	357	338	444	286	430	366	210	251
Similarity (%)	79.6				72.1	58.5	54.9	71.2	76.5	56.5	51.4	68.6	59.2	76.8	58.4	85.8	73.0	85.7	84.5
Identity (%)	55.4				45.5	24.2	26.2	42.9	56.4	24.6	26.0	39.5	32.8	40.8	28.0	59.8	45.6	63.3	60.2
Homologous gene	Streptomyces coelicolor A3(2) folC				Bacillus subtilis 168 balS	Bacillus subtilis 168 oppA	Bacillus subtilis 168 dnaK	Eikenella corrodens ATCC 23824	Thermus aquaticus ATCC 33923 mdh	Streptomyces coelicolor A3(2) SC4A10.33	Vibrio cholerae aphA	Acinetobacter sp. vanA	Sphingomonas flava ATCC 39723 pcpD	Acinetobacter sp. vanK	Klebsiella pneumoniae mdcF	Bacillus subtilis clpX	Streptomyces coelicolor A3(2) SCF55.28c	Streptomyces sp. 2065 pcaJ	Streptomyces sp. 2065 pcal
db Match	prf.2410252B	·			sp:SYV_BACSU	pir.A38447	sp:DNAK_BACSU	gp:ECU89166_1	sp:MDH_THEFL	gp:SC4A10_33	gp:AF065442_1	prf:2513416F	gp:FSU12290_2	prf.2513416G	gp:KPU95087_7	prf:2303274A	gp:SCF55_28	gp:AF109386_2	gp:AF109386_1
ORF (bp)	1374	612	714	663	2700	1575	1452	585	984	777	576	1128	975	1425	930	1278	1086	633	750
Terminal (nt)	2514114	2516273	2516956	2517751	2515637	2518398	2521660	2521667	2522265	2524337	2524340	2526226	2527207	2528559	2528551	2529484	2531976	2531969	2532604
Initial (nt)	2515487	2515662	2516243	2517089	2518336	2519972	2520209	2522251	2523248	2523561	2524915	2525099	2526233	2527135	2529480	2530761	2530891	2532601	2533353
SEQ NO. (a.a.)	6104	6105	6106	6107	6108	6109	6110	6111	6112	6113	6114	6115	6116	6117	6118	6119	6120	6121	6122
SEQ NO. (DNA)	2604	2605	2606	2607	2608	2609	2610	2611	2612	2613	2614	2615	2616	2617	2618	2619	2620	2621	2622

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

	Function	protocatechuate catabolic protein	beta-ketothiolase		3-oxoadipate enol-lactone hydrolase and 4-carboxymuconolactone decarboxylase	transcriptional regulator	3-oxoadipate enol-lactone hydrolase and 4-carboxymuconolactone decarboxylase		3-carboxy-cis, cis-muconate cycloisomerase	protocatechuate dioxygenase alpha subunit	protocatechuate dioxygenase beta subunit	hypothetical protein	muconolactone isomerase		muconate cycloisomerase		catechol 1,2-dioxygenase		toluate 1,2 dioxygenase subunit
-	Matched length (a.a.)	251	406		256	825	115		437	214	217	273	92		372		285		437
	Similarity (%)	82.5	71.9		76.6	43.0	89.6		63.4	70.6	91.2	48.7	81.5		84.7		88.4		85.6
	Identity (%)	58.2	44.8		50.8	23.6	78.3		39.8	49.5	74.7	26.4	54.4		60.8		72.3		62.2
Table 1 (continued)	Homologous gene	Rhodococcus opacus 1CP pcaR	Ralstonia eutropha bktB		Rhodococcus opacus pcal.	Streptomyces coelicolor A3(2) SCM1.10	Rhodococcus opacus pcal.		Rhodococcus opacus pcaB	Rhodococcus opacus pcaG	Rhodococcus opacus pcaH	Mycobacterium tuberculosis H37Rv Rv0336	Mycobacterium tuberculosis catC		Rhodococcus opacus 1CP catB		Rhodococcus rhodochrous catA		Pseudomonas putida plasmid pDK1 xylX
	db Match	prf:2408324F	prf:2411305D		prf:2408324E	gp:SCM1_10	prf:2408324E		prf.2408324D	prf:2408324C	prf:2408324B	pir.G70506	prf:2515333B		sp:CATB_RHOOP	-	prf:2503218A		gp.AF134348_1
	ORF (bp)	792	1224	912	753	2061	366	678	1116	612	069	1164	291	771	1119	909	855	141	1470
	Terminal (nt)	2534182	2535424	2534257	2536182	2538256	2538248	2540230	2538616	2539709	2540335	2541187	2542512	2543813	2542818	2544867	2544022	2544928	2546784
	Initial (nt)	2533391	2534201	2535168	2535430	2536196	2538613	2539553	2539731	2540320	2541024	2542350	2542802	2543043	2543936	2544262	2544876	2545068	6140 2545315
	SEQ NO. (a.a.)	6123	6124	6125	6126	6127	6128	6129	6130	6131	6132	6133	6134	6135	6136	6137	6138	6139	6140
	SEQ NO. (DNA)	2623	2624	2625	2626	2627	2628	2629	2630	2631	2632	2633	2634	2635	2636	2637	2638	2639	2640

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

$\overline{}$
רד
ŏ
w
_
=
=
ᆮ
8
ပ
_
$\overline{}$
യ
<u> </u>
Table
m
$\overline{}$

_																			
	Function	toluate 1,2 dioxygenase subunit	toluate 1,2 dioxygenase subunit	1,2-dihydroxycyclohexa-3,5-diene carboxylate dehydrogenase	regulator of LuxR family with ATP-binding site	transmembrane transport protein or 4-hydroxybenzoate transporter	benzoate membrane transport protein	ATP-dependent Clp protease proteolytic subunit 2	ATP-dependent Clp protease proteolytic subunit 1	hypothetical protein	trigger factor (prolyl isomerase) (chaperone protein)	hypothetical protein	penicillin-binding protein	hypothetical protein		transposase		hypothetical protein	transposase
	Matched length (a.a.)	161	342	277	979	435	388	197	198	42	417	160	336	115		142		35	75
	Similarity (%)	83.2	81.0	61.4	48.6	64.4	66.2	88.3	85.9	71.4	66.4	63.1	50.9	58.3		73.2		82.9	78.7
	Identity (%)	60.3	51.5	30.7	23.3	31.3	29.9	69.5	62.1	42.9	32.1	32.5	25.3	27.8		54.2		57.1	50.7
	Homologous gene	Pseudomonas putida plasmid pDK1 xylY	Pseudomonas putida plasmid pDK1 xylZ	Pseudomonas putida plasmid pDK1 xylL	Rhodococcus erythropolis thcG	Acinetobacter calcoaceticus pcaK	Acinetobacter calcoaceticus benE	Streptomyces coelicolor M145 clpP2	Streptomyces coelicolor M145 clpP1	Sulfolobus islandicus ORF154	Bacillus subtilis 168 tig	Streptomyces coelicolor A3(2) SCD25.17	Nocardia lactamdurans LC411 pbp	Mus musculus Moa1		Corynebacterium striatum ORF1		Corynebacterium striatum ORF1	Corynebacterium striatum ORF1
	db Match	gp:AF134348_2	gp:AF134348_3	gp:AF134348_4	gp:REU95170_1	sp:PCAK_ACICA	sp:BENE_ACICA	gp:AF071885_2	gp:AF071885_1	gp:SIS243537_4	sp:TIG_BACSU	gp:SCD25_17	sp:PBP4_NOCLA	prf:2301342A		prf.2513302C		prf:2513302C	prf.2513302C
	ORF (bp)	492	1536	828	2685	1380	1242	624	603	150	1347	495	975	456	249	438	150	126	264
	Terminal (nt)	2547318	2548868	2549695	2552455	2553942	2555267	2555317	2555978	2556748	2556760	2559103	2560131	2560586	2561363	2561483	2562242	2561990	2562078
	Initial (nt)	2546827	2547333	2548868	2549771	2552563	2554026	2555940	2556580	2556599	2558106	2558609	2559157	2560131	2561115	2561920	2562093	2562115	2562341
	SEQ NO. (a.a.)	6141	6142	6143	6144	6145	6146	6147	6148	6149	6150	6151	6152	6153	6154	6155	6156	6157	6158
	SEQ NO.	2641	2642	2643	2644	2645	2646	2647	2648	2649	2650	2651	2652	2653	2654	2655	2656	2657	2658

									_			_										
10	Function			galactose-6-phosphate isomerase	hypothetical protein	hypothetical protein	aminopeptidase N	hypothetical protein				phytoene desaturase			phytoene dehydrogenase	phytoene synthase	multidrug resistance transporter		ABC transporter ATP-binding protein	dipeptide transport system permease protein	nickel transport system permease protein	
15	Matched length (a.a.)			140	248	199	890	358				104			381	290	392		538	286	316	
20	Similarity (%)			71.4	58.1	80.9	70.5	58.1				81.7			63.8	58.6	47.7		71.6	73.8	62.0	
	Identity (%)			40.0	26.2	56.8	47.5	25.1				61.5			31.2	31.4	25.8		41.3	38.8	33.2	
os fable 1 (continued)	ns gene			reus NCTC	ulyticus ORF2	berculosis	ans pepN	ri BB0852				ens ATCC			nus DK1050	eus JA3933	genes IItB	,	ongatus	4 dppC	12 nikB	
39 Table 1 ((	Homologous gene			Staphylococcus aureus NCTC 8325-4 lacB	Bacillus acidopullulyticus ORF2	Mycobacterium tuberculosis H37Rv Rv2466c	Streptomyces lividans pepN	Borrelia burgdorferi BB0852				Brevibacterium linens ATCC 9175 crtl			Myxococcus xanthus DK1050 carA2	Streptomyces griseus JA3933 crtB	Listeria monocytogenes IItB		Synechococcus elongatus	Bacillus firmus OF4 dppC	Escherichia coli K12 nikB	
35						· 2 T		В											2		Ш	
40	db Match			sp:LACB_STAAU	sp:YAMY_BACAD	pir:A70866	SP. AMPN_STRLI	pir:B70206				gp:AF139916_3			sp.CRTJ_MYXXA	sp:CRTB_STRGR	gp:LMAJ9627_3		gp:SYOATPBP	sp.DPPC_BACFI	pir:S47696	
	ORF (bp)	390	885	471	969	609	2601	1083	1152	999	156	327	171	378	1206	876	1119	1233	1641	882	939	1707
45	Terminal (nt)	2562387	2563847	2563932	2564550	2565623	2568945	2570293	2570309	2572175	2572348	2572351	2572807	2573393	2572659	2573843	2574780	2575981	2577232	2578879	2579769	2580711
50	Initial (nt)	2562776	2562963	2564402	2565245	2566231	2566345	2569211	2571460	2571510	2572193	2572677	2572977	2573770	2573864	2574718	2575898	2577213	2578872	2579760	2580707	2582417
	SEQ NO. (a.a.)	6159	6160	6161	6162	6163	6164	6165	6166	6167	6168	6169	6170	6171	6172	6173	6174	6175	6176	6177	6178	6179
55	SEO NO (DNA)	2659	2660	2661	2662	2663	2664	2665	2666	2667	2668	2669	2670	2671	2672	2673	2674	2675	2676	2677	2678	2679

5		Function		acetylornithine aminotransferase	hypothetical protein	hypothetical membrane protein	acetoacetyl CoA reductase	transcriptional regulator, TetR family	polypeptides predicted to be useful antigens for vaccines and diagnostics	ABC transporter ATP-binding protein	globin	chromate transport protein	hypothetical protein	hypothetical protein		hypothetical protein	ABC transporter ATP-binding protein	hypothetical protein	hypothetical membrane protein	alkaline phosphatase
15		Matched length (a.a.)		411 ac	482 hy	218 h)	235 ac	240 tr	94 di a D	238 A	126 gl	396 cł	196 h	127 h		55 h	563 A	172 h	700 h	536 al
20		Similarity (%)		63.5	47.9	79.4	0.09	55.0	47.0	65.1	77.0	60.4	68.9	61.4		0.09	79.6	62.2	56.7	52.6
		Identity (%)		31.4	25.1	49.1	28.1	26.7	38.0	31.1	53.2	27.3	37.8	36.2		36.4	52.8	31.4	28.0	28.0
25	ntinued)	gene		ıtamicum	rculosis	rculosis	n D phbB	olor actil	lis	a GM73	ае	ginosa ırA	erculosis	olor A3(2)		1 APE1182	2 yijK	erculosis	ае об59	8
30	Table 1 (continued)	Homologous gene		Corynebacterium glutamicum ATCC 13032 argD	Mycobacterium tuberculosis H37Rv Rv1128c	Mycobacterium tuberculosis H37Rv Rv0364	Chromatium vinosum D phbB	Streptomyces coelicolor actll	Neisseria meningitidis	Pseudomonas putida GM73 ttg2A	Mycobacterium leprae MLCB1610.14c	Pseudomonas aeruginosa Plasmid pUM505 chrA	Mycobacterium tuberculosis H37Rv Rv2474c	Streptomyces coelicolor A3(2) SC6D10.19c		Aeropyrum pernix K1 APE1182	Escherichia coli K12 yjjK	Mycobacterium tuberculosis H37Rv Rv2478c	Mycobacterium leprae o659	Bacillus subtilis phoB
40		db Match		sp.ARGD_CORGL	pir.A70539	sp:YA26_MYCTU	sp:PHBB_CHRVI (	pir.A40046	75	gp:AF106002_1	gp:MLCB1610_9	sp.CHRA_PSEAE	pir.A70867	gp:SC6D10_19		pir.B72589	sp:YJJK_ECOLI	pir.E70867	sp:Y05L_MYCLE	pir.C69676 E
		ORF (bp)	1941	1314	1584	747	708	738	441	792	393	1128	627	465	621	162	1668	615	2103	1419
45		Terminal (nt)	2584504	2585926	2587763	2588722	2588725	2590302	2591137	2591574	2592794	2593965	2593968	2594597	2595188	2595822	2596048	2597869	2598662	2602879
50		Initial (nt)	2582564	2584613	2586180	2587976	2589432	2589565		2592365	2592402	2592838	2594594	2595061	2595808	2595983	2597715	2598483	2600764	2601461
		SEQ NO.	6180	6181	6182	6183	6184	6185	6186	6187	6188	6189	6190	6191	6192	6193	6194	6195	6196	6197
55		SEQ NO (DNA)	2680	2681	2682	2683	2684	2685	2686	2687	2688	2689	2690	2691	2692	2693	2694	2695	2696	2697

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

J	
10	
15	
20	
25	
30	
35	
40	
45	
50	

Table 1 (continued)

														_				
Function			multiple sugar-binding transport system permease protein	multiple sugar-binding transport system permease protein		maltose-binding protein		ABC transporter ATP-binding protein (ABC-type sugar transport protein) or cellobiose/maltose transport protein		dolichol phosphate mannose synthase		aldehyde dehydrogenase	circadian phase modifier		hypothetical membrane protein	glyoxylate-induced protein	ketoacyl reductase	oligoribonuclease
Matched length (a.a.)			279	292		462		386		154		207	183		412	255	258	179
Similarity (%)			76.3	67.5		63.2		79.8		72.7.		89.4	73.8		64.6	69.4	57.0	78.8
Identity (%)			39.1	27.4		28.8		59.1		37.7		67.2	48.6		35.0	41.2	40.0	48.0
Homologous gene			Streptococcus mutans INGBRITT msmG	Streptococcus mutans INGBRITT msmF		Thermoanaerobacterium thermosul amyE		Streptomyces reticuli msiK		Schizosaccharomyces pombe dpm1		Rhodococcus rhodochrous plasmid pRTL1 orf5	Synechococcus sp. PCC7942 cpmA		Thermotoga maritima MSB8 TM0964	Escherichia coli K12 gip	Mycobacterium tuberculosis H37Rv Rv1544	Escherichia coli K12 orn
db Match			sp:MSMG_STRMU	sp:MSMF_STRMU		prf:2206392C		prf.2308356A		prf.2317468A		prf.2516398E	prf.2513418A		pir:A72312	sp:GIP_ECOLI	pir.E70761	sp:ORN_ECOLI
ORF (bp)	930	639	912	843	1674	1329	1242	1128	750	684	069	789	762	345	1182	750	798	657
Terminal (nt)	2605502	2603945	2604609	2605527	2608117	2606561	2608185	2609512	2612272	2610848	2613151	2614500	2615410	2615795	2615939	2617995	2618869	2619538
Initial (nt)	2604573	2604583	2605520	2606369	2606444	2607889	2609426	2610639	2611523	2611531	2612462	2613712	2614649	2615451	2617120	2617246	2618072	2618882
SEQ NO. (a.a.)	6198	6199	6200	6201	6202	6203	6204	6205	6206	6207	6208	6209	6210	6211	6212	6213	6214	6215
SEQ NO. (DNA)	2698	2699	2700	2701	2702	2703	2704	2705	2706	2707	2708	2709	2710	2711	2712	2713	2714	2715

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

	Function	ferric enterochelin esterase	lipoprotein				transposase (IS1207)			transcriptional regulator	glutaminase	sporulation-specific degradation regulator protein		uronate isomerase		hypothetical protein	pyrazinamidase/nicotinamidase	hypothetical protein	bacterioferritin comigratory protein	bacterial regulatory protein, tetR family
	Matched length (a.a.)	454 fer	398 lipo				436 tra			131 tra	358 glt	gs 76		335 ur		291 hy	185 py	75 hy	141 ba	114 ba
	Similarity (%)	50.9	71.9				8.66			63.4	69.3	72.2		6.09		45.0	74.6	80.0	73.8	61.4
	Identity (%)	26.0	48.5				99.5			32.8	35.2	42.3		29.0		32.0	48.1	42.7	46.8	32.5
Table 1 (continued)	Homologous gene	Salmonella enterica iroD	Mycobacterium tuberculosis H37Rv Rv2518c lppS				Corynebacterium glutamicum ATCC 21086			Salmonella typhimurium KP1001 cytR	Rattus norvegicus SPRAGUE. DAWLEY KIDNEY	Bacillus subtilis 168 degA		Escherichia coli K12 uxaC		Zea diploperennis perennial teosinte	Mycobacterium avium pncA	Mycobacterium tuberculosis H37Rv Rv2520c	Escherichia coli K12 bcp	Streptomyces coelicalor A3(2) SCI11.01c
	db Match	prf:2409378A	pir:C70870				gp:SCU53587_1			gp:AF085239_1	sp:GLSK_RAT	pir.A36940		sp:UXAC_ECOLI		prf:1814452C	prf:232444A	pir.E70870	sp:BCP_ECOLI	gp:SCI11_1
	ORF (bp)	1188	1209	645	150	246	1308	207	639	453	1629	477	555	1554	501	1197	558	273	465	636
	Terminal (nt)	2619541	2620973	2623605	2623621	2624048	2624051	2625806	2625809	2628376	2626493	2628852	2628324	2630479	2631136	2632466	2633100	2633146	2634064	2634751
	Initial (nt)	2620728	2622181	2622961	2623770	2623803	2625358	2625600	2626447	2627924	2628121	2628376	2628878	2628926	2630636	2631270	2632543	2633418	2633600	6234 2634116
	SEQ NO.	6216	6217	6218	6219	6220	6221	6222	6223	6224	6225	6226	6227	6228	6229	6230	6231	6232	6233	
	SEQ NO.	2716	2717	2718	2719	2720	2721	2722	2723	2724	2725	2726	2727	2728	2729	2730	2731	2732	2733	2734

5	
10	
15	
20	•
25	
30	
35	
40	
45	
50	

Table 1 (continued)	Function	phosphopantethiene protein transferase	lincomycin resistance protein	hypothetical membrane protein		fatty-acid synthase	hypothetical protein	peptidase	hypothetical membrane protein	hypothetical membrane protein	hypothetical protein	ribonuclease PH				hypothetical membrane protein	transposase (IS1628)		aryisulfatase
	Matched length (a.a.)	145	473	113		3029	404	230	112	113	202	236				428	175		250
	Similarity (%)	75.9	85.6	54.0		83.6	55.2	6.09	6'29	0'69	2.92	81.4				58.2	97.2		74.4
	Identity (%)	56.6	52.4	30.1		62.3	25.3	40.4	40.2	37.2	55.0	60.2				29.0	92.1		46.0
	Homalogous gene	Corynebacterium ammoniagenes ATCC 6871 ppt1	Corynebacterium glutamicum ImrB	Synechocystis sp. PCC6803		Corynebacterium ammoniagenes fas	Streptomyces coelicolor A3(2) SC4A7.14	Mycobacterium tuberculosis H37Rv Rv0950c	Mycobacterium tuberculosis H37Rv Rv1343c	Mycobacterium leprae B1549_F2_59	Mycobacterium tuberculosis H37Rv Rv1341	Pseudomonas aeruginosa ATCC 15692 rph				Mycobacterium tuberculosis H37Rv SC8A6.09c	Corynebacterium glutamicum 22243 R-plasmid pAG1 tnpB		Mycobacterium leprae ats
	db Match	gp:BAY15081_1	gp:AF237667_1	pir.S76537		pir:S2047	gp:SC4A7_14	pir:D70716	sp:Y077_MYCT	sp:Y076_MYCLE	sp:Y03Q_MYCTU	sp:RNPH_PSEAE				sp:Y029_MYCTU	gp:AF121000_8		sp:Y03O_MYCLE
	ORF (bp)	405	1425	324	414	8979	1182	615	462	354	618	735	246	693	582	1362	534	099	765
İ	Terminal (nt)	2634747	2635165	2637168	2637240	2638649	2648235	2650164	2650902	2651339	2651420	2652067	2653009	2653326	2654079	2654875	2656985	2656974	2657736
	Initial (nt)	2635151	2636589	2636845	2637653	2647627	2649416	2649550	2650441	2650986	2652037	2652801	2653254	2654018	2654660	2656236	2656452	2657633	2658500
	SEQ NO. (a.a.)	6235	6236	6237	6238	6239	6240	6241	6242	6243	6244	6245	6246	6247	6248	6249	6250	6251	6252
	SEQ NO. (DNA)	2735	2736	2737	2738	2739	2740	2741	2742	2743	2744	2745	2746	2747	2748	2749	2750	2751	2752

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

_																			
	Function	D-glutamate racemase	·	bacterial regulatory protein, marR family	hypothetical membrane protein		endo-type 6-aminohexanoate oligomer hydrolase	hypothetical protein	hypothetical protein		hypothetical protein		ATP-dependent helicase	hypothetical membrane protein	hypothetical protein	phosphoserine phosphatase		cytochrome c oxidase chain I	
	Matched length (a.a.)	284		147	225		321	200	105		428		647	313	222	310		575	
	Similarity (%)	99.3		70.8	69.3		58.3	58.5	77.1		80.8		53.3	60.1	52.0	61.0		74.4	
]	Identity (%)	99.3		44.2	38.2		30.2	35.0	57.1		61.2		25.2	29.7	39.0	38.7		46.8	
	Homologous gene	Corynebacterium glutamicum ATCC 13869 murl		Streptomyces coelicolor A3(2) SCE22.22	Mycobacterium tuberculosis H37Rv Rv1337		Flavobacterium sp. nylC	Mycobacterium tuberculosis H37Rv Rv1332	Mycobacterium tuberculosis H37Rv Rv1331		Mycobacterium tuberculosis H37Rv Rv1330c		Escherichia coli dinG	Mycobacterium tuberculosis H37Rv Rv2560	Streptomyces coelicolor A3(2) SC1B5.06c	Escherichia coli K12 serB		Mycobacterium tuberculosis H37Rv Rv3043c	
	db Match	prf:2516259A		gp:SCE22_22	sp:Y03M_MYCTU	-	pir.A47039	sp:Y03H_MYCTU	sp:Y03G_MYCTU		sp:Y03F_MYCTU		prf. 1816252A	sp:Y0A8_MYCTU	pir.T34684	sp:SERB_ECOLI		pir:D45335	
ĺ	ORF (bp)	852	636	492	747	891	096	537	300	624	1338	306	1740	891	723	1017	1596	1743	306
	Terminal (nt)	2658606	2660131	2660147	2660671	2662455	2661417	2662331	2662883	2664060	2665397	2665992	2667854	2667870	2668839	2669557	2672721	2671063	2673255
	Initial (nt)	2659457	2659496	2660638	2661417	2661565	2662376	2662867	2663182	2663437	2664060	2665687	2666115	2668760	2669561	2670573	2671126	2672805	2672950
	SEQ NO. (a.a.)	6253	6254	6255	6256	6257	6258	6259	6260	6261	6262	6263	6264	6265	6266	6267	6268	6569	6270
	SEQ NO. (DNA)	2753	2754	2755	2756	2757	2758	2759	2760	2761	2762	2763	2764	2765	2766	2767	2768	2769	2770

5	
10	
15	•
20	
25	
30	
35	
40	
45	
50	

|                     | Function                    | ribonucleotide reductase beta-chain                                       | ferritin   
   
   | sporulation transcription factor   | iron dependent repressor or<br>diptheria toxin repressor | cold shock protein TIR2 precursor                   | hypothetical membrane protein  | ribonucleotide reductase alpha-<br>chain  |  
   | 50S ribosomal protein L36   
  | NH3-dependent NAD(+) synthetase  |   |   | hypothetical protein  | hypothetical protein   | alcohol dehydrogenase  
  | Bacillus subtilis mmg (for mother cell<br>metabolic genes)  | hypothetical protein          |         | phosphoglucomutase       |
|---------------------|-----------------------------|---
--
--
--|--|--|---|--|---
--
--|--|--|---
---|---|--|---|---|-------------------------------|---------|--------------------------|
|                     | Matched<br>length<br>(a.a.) | 334   | 159  
   
   | 256  | 225  | 124   | 50   | 707   |  
   | 41  
  | 279  |   |   | 257   | 96   | 337  
  | 459   | 284                           |         | 556                      |
|                     | Similarity<br>(%)           | 99.7  | 64.2   
   
   | 60.2   | 60.4   | 62.1  | 0.98   | 100.0   |  
   | 79.0  
  | 78.1   |   |   | 56.4  | 68.8   | 52.8   
  | 56.0  | 66.2                          |         | 90.6                     |
|                     | Identity<br>(%)             | 99.7  | 31.5   
   
   | 32.8   | 27.6   | 24.2  | 50.0   | 6.99  |  
   | 58.0  
  | 55.6   |   |   | 30.7  | 41.7   | 26.1   
  | 27.0  | 33.8                          |         | 61.7                     |
| Table 1 (continued) | Homologous gene             | Corynebacterium glutamicum<br>ATCC 13032 nrdF                             | Escherichia coli K12 ftnA  
   
   | Streptomyces coelicolor A3(2) whiH   | Corynebacterium glutamicum<br>ATCC 13869 dtxR            | Saccharomyces cerevisiae<br>YPH148 YOR010C TIR2     | Archaeoglobus fulgidus AF0251  | Corynebacterium glutamicum<br>ATCC 13032 nrdE   |  
   | Rickettsia prowazekii   
  | Bacillus subtilis 168 nadE   |   |   | Synechacystis sp. PCC6803<br>slr1563  | Mycobacterium tuberculosis<br>H37Rv Rv3129   | Bacillus stearothermophilus<br>DSM 2334 adh  
  | Bacillus subtilis 168 mmgE  | Arabidopsis thaliana T6K22.50 |         | Escherichia coli K12 pgm |
|                     | db Match                    | gp:AF112536_1   | sp:FTNA_ECOLI  
   
   | gp:SCA32WHIH_4   | pir.140339   | sp:TIR2_YEAST                                       | pir.C69281   | gp:AF112535_3   |  
   | SP:RL36_RICPR   
  | sp:NADE_BACSU  |   |   | pir.S76790  | pir.G70922   | sp:ADH2_BACST  
  | sp:MMGE_BACSU   | pir.T05174                    |         | sp:PGMU_ECOLI            |
|                     | ORF<br>(bp)                 | 1002  | 486  
   
   | 750  | 099  | 438   | 276  | 2121  | 315  
   | 141   
  | 831  | 93  | 498   | 747   | 288  | 1020   
  | 1371  | 834                           | 792     | 1662                     |
|                     | Terminal<br>(nt)            | 2673338   | 2675289  
   
   | 2676240  | 2676243  | 2677377   | 2676918  | 2677478   | 2680784  
   | 2681223   
  | 2682376  | 2681464   | 2683616   | 2682379   | 2683131  | 2683627  
  | 2686289   | 2687148                       | 2687449 | 2688389                  |
|                     | Initial<br>(nt)             | 2674339   | 2674804  
   
   | 2675491  | 2676902  | 2676940   | 2677193  | 2679598   | 2680470  
   | 2681363   
  | 2681546  | 2681556   | 2683119   | 2683125   | 2683418  | 2684646  
  | 2684919   | 2686315                       | 2688240 | 2690050                  |
|                     |                             | 6271  | 6272   
   
   | 6273   | 6274   | 6275  | 6276   | 6277  | 6278   
   |   
  | <del>-</del> -   | 6281  | 6282  | 6283  | 6284   |  
  | 6286  | 6287                          |         | 6289                     |
|                     | SEQ<br>NO.                  | 2771  | 2772   
   
   | 2773   | 2774   | 2775  | 2776   | 2777  | 2778   
   | 2779  
  | 2780   | 2781  | 2782  | 2783  | 2784   | 2785   
  | 2786  | 2787                          | 2788    | 2789                     |
|                     | Table 1 (continued)         | SEQ Initial Terminal ORF db Match Homologous gene (%) (nt) (nt) (bp) (bp) | SEQ Initial (a.a.)         Terminal (ht)         ORF (ht)         db Match         Homologous gene (ht)         Identity (ht)         Similarity length (ht)         Matched (ht)           (a.a.)         (nt)         (nt)         (ht)         (ht) </td <td>SEQ Initial (a.a.)         Terminal (a.a.)         Terminal (bp)         db Match (bp)         Homologous gene (a.a.)         Identity (bp)         Matched (a.a.)           6271         2674339         2673338         1002         gp:AF112536_1         Corynebacterium glutamicum g</td> <td>SEQ (nt) (nt) (nt) (nt) (nt) (nt) (nt) (nt)</td> <td>SEQ Initial (nt) (nt) (nt) (nt) (nt) (nt) (nt) (nt)</td> <td>SEQ (nt)         Initial (nt)         Terminal (nt)         ORF (bp)         db Match         Homologous gene (%)         Identity (%)         Matched (%)</td> <td>SEQ (nt)         Initial (nt)         Terminal (nt)         ORF (bp)         db Match         Homologous gene (%)         Identity (%)         Similarity (%)         Matched (%)           6271         2674339         2673338         1002         gp.AF112536_1         ATCC 13032 nrdF         99.7         334           6272         2674804         2675289         486         sp.FTNA_ECOLI         Escherichia coli K12 ftnA         31.5         64.2         159           6273         2675491         2676240         750         gp.SCA32WHIH_A         Streptomyces coelicolor A3(2)         32.8         60.2         256           6274         2676902         2676243         660         pir.140339         ATCC 13869 dtxR         27.6         60.4         225           6275         2676940         2677377         438         sp.TIR2_YEAST         YPH148 YOR010C TIR2         24.2         62.1         124           6276         2677193         2676918         276         pir.C69281         Archaeoglobus fulgidus AF0251         50.0         86.0         50</td> <td>SEQ         Initial         Terminal         ORF         db Match         Homologous gene         Identity (%)         Similarity length (%)         Matched (%)           NO. (nt)         (nt)         (nt)         (bp)         db Match         Homologous gene         (%)<td>SEQ (a.a.)         Initial (bt)         Terminal (bp)         ORF (bp)         db Match         Homologous gene (%)         Identity (%)         Similarity (hatched)           6271         2674339         2673338         1002         gp:AF112536_1         Corynebacterium glutamicum (%)         99.7         334           6272         2674804         2675289         486         sp:FTNA_ECOLI         Escherichia coli K12 ftnA         31.5         64.2         159           6273         2675491         2676240         750         gp:SCA32WHIH_4         Streptomyces coelicolor A3(2)         32.8         60.2         256           6274         2676940         2676243         660         pir.140339         Corynebacterium glutamicum (plutamicum 27.6         60.4         225           6275         2676940         2677377         438         sp:TIR2_YEAST         Saccharomyces crewisiae         24.2         62.1         124           6276         2677193         2676918         276         pir.C69281         Archaeoglobus fulgidus AF0251         50.0         86.0         50           6277         2679588         2677478         2121         gp:AF112535_3         Corynebacterium glutamicum 399.9         100.0         707</td><td>  SEQ   Initial   Terminal   ORF   db Match   Homologous gene   (%)   (%</td><td>SEQ Initial (a.a.)         Terminal (bp)         CRF (bp)         db Match (bp)         Homologous gene (bp)         Identity (bp)         Matched (bp)           8271 2674339 (a.a.)         2673338 (bp)         1002 gp:AF112536_1         Corynebacterium glutamicum gg.7 (bp)         99.7 (bp)         334           6272 2674804 2675289 (bp)         2675289 486 sp:FTNA_ECOLI (bp)         Escherichia coli K12 ftnA (bp)         31.5 (bp)         60.2 (bp)         256           6273 2676902 2676243 (bp)         2676249 750 gp:SCA32WHIH_A (bp)         Streptomyces coelicolor A3(2) 32.8 (bp)         32.8 (bp)         256           6276 2676940 2676243 (bp)         2676243 (bp)         2676243 (bp)         2676344 (bp)         27.6 (bp)         27.6 (bp)         27.6 (bp)           6276 2676940 2677377 (bp)         27.6 (bp)         27.7 (bp)         27.7 (bp)         27.7 (bp)         27.7 (bp)         27.7 (bp)           6276 277 2676948 276 (bp)         27.7 (bp)</td><td>SEQ         Initial         Terminal         ORF         db Match         Homologous gene         Identity (%)         Similarity (%)         Matched (%)           NOD (11)         (I1)         (ID)         &lt;</td><td>  Table 1 (continued)   SEQ   Initial   Terminal   CRF   Ab Match   Homologous gene   (%)</td><td>  Table 1 (continued)   Terminal   ChPF   Ab Match   Homologous gene   Homologous gene   (%)   (</td><td>  Table 1 (continued)   SEQ   Initial   Terminal   ORF   db Match   Homologous gene   (%)</td><td>SEQ         Initial         Teminal         ORF         db Match         Homologous gene         (%)         Matched           (ND)         (Int)         (Int)</td><td>  SEQ</td><td>  SEG</td><td>  SEG</td></td> | SEQ Initial (a.a.)         Terminal (a.a.)         Terminal (bp)         db Match (bp)         Homologous gene (a.a.)         Identity (bp)         Matched (a.a.)           6271         2674339         2673338         1002         gp:AF112536_1         Corynebacterium glutamicum g | SEQ (nt) (nt) (nt) (nt) (nt) (nt) (nt) (nt)              | SEQ Initial (nt) (nt) (nt) (nt) (nt) (nt) (nt) (nt) | SEQ (nt)         Initial (nt)         Terminal (nt)         ORF (bp)         db Match         Homologous gene (%)         Identity (%)         Matched (%) | SEQ (nt)         Initial (nt)         Terminal (nt)         ORF (bp)         db Match         Homologous gene (%)         Identity (%)         Similarity (%)         Matched (%)           6271         2674339         2673338         1002         gp.AF112536_1         ATCC 13032 nrdF         99.7         334           6272         2674804         2675289         486         sp.FTNA_ECOLI         Escherichia coli K12 ftnA         31.5         64.2         159           6273         2675491         2676240         750         gp.SCA32WHIH_A         Streptomyces coelicolor A3(2)         32.8         60.2         256           6274         2676902         2676243         660         pir.140339         ATCC 13869 dtxR         27.6         60.4         225           6275         2676940         2677377         438         sp.TIR2_YEAST         YPH148 YOR010C TIR2         24.2         62.1         124           6276         2677193         2676918         276         pir.C69281         Archaeoglobus fulgidus AF0251         50.0         86.0         50 | SEQ         Initial         Terminal         ORF         db Match         Homologous gene         Identity (%)         Similarity length (%)         Matched (%)           NO. (nt)         (nt)         (nt)         (bp)         db Match         Homologous gene         (%) <td>SEQ (a.a.)         Initial (bt)         Terminal (bp)         ORF (bp)         db Match         Homologous gene (%)         Identity (%)         Similarity (hatched)           6271         2674339         2673338         1002         gp:AF112536_1         Corynebacterium glutamicum (%)         99.7         334           6272         2674804         2675289         486         sp:FTNA_ECOLI         Escherichia coli K12 ftnA         31.5         64.2         159           6273         2675491         2676240         750         gp:SCA32WHIH_4         Streptomyces coelicolor A3(2)         32.8         60.2         256           6274         2676940         2676243         660         pir.140339         Corynebacterium glutamicum (plutamicum 27.6         60.4         225           6275         2676940         2677377         438         sp:TIR2_YEAST         Saccharomyces crewisiae         24.2         62.1         124           6276         2677193         2676918         276         pir.C69281         Archaeoglobus fulgidus AF0251         50.0         86.0         50           6277         2679588         2677478         2121         gp:AF112535_3         Corynebacterium glutamicum 399.9         100.0         707</td> <td>  SEQ   Initial   Terminal   ORF   db Match   Homologous gene   (%)   (%</td> <td>SEQ Initial (a.a.)         Terminal (bp)         CRF (bp)         db Match (bp)         Homologous gene (bp)         Identity (bp)         Matched (bp)           8271 2674339 (a.a.)         2673338 (bp)         1002 gp:AF112536_1         Corynebacterium glutamicum gg.7 (bp)         99.7 (bp)         334           6272 2674804 2675289 (bp)         2675289 486 sp:FTNA_ECOLI (bp)         Escherichia coli K12 ftnA (bp)         31.5 (bp)         60.2 (bp)         256           6273 2676902 2676243 (bp)         2676249 750 gp:SCA32WHIH_A (bp)         Streptomyces coelicolor A3(2) 32.8 (bp)         32.8 (bp)         256           6276 2676940 2676243 (bp)         2676243 (bp)         2676243 (bp)         2676344 (bp)         27.6 (bp)         27.6 (bp)         27.6 (bp)           6276 2676940 2677377 (bp)         27.6 (bp)         27.7 (bp)         27.7 (bp)         27.7 (bp)         27.7 (bp)         27.7 (bp)           6276 277 2676948 276 (bp)         27.7 (bp)</td> <td>SEQ         Initial         Terminal         ORF         db Match         Homologous gene         Identity (%)         Similarity (%)         Matched (%)           NOD (11)         (I1)         (ID)         &lt;</td> <td>  Table 1 (continued)   SEQ   Initial   Terminal   CRF   Ab Match   Homologous gene   (%)</td> <td>  Table 1 (continued)   Terminal   ChPF   Ab Match   Homologous gene   Homologous gene   (%)   (</td> <td>  Table 1 (continued)   SEQ   Initial   Terminal   ORF   db Match   Homologous gene   (%)</td> <td>SEQ         Initial         Teminal         ORF         db Match         Homologous gene         (%)         Matched           (ND)         (Int)         (Int)</td> <td>  SEQ</td> <td>  SEG</td> <td>  SEG</td> | SEQ (a.a.)         Initial (bt)         Terminal (bp)         ORF (bp)         db Match         Homologous gene (%)         Identity (%)         Similarity (hatched)           6271         2674339         2673338         1002         gp:AF112536_1         Corynebacterium glutamicum (%)         99.7         334           6272         2674804         2675289         486         sp:FTNA_ECOLI         Escherichia coli K12 ftnA         31.5         64.2         159           6273         2675491         2676240         750         gp:SCA32WHIH_4         Streptomyces coelicolor A3(2)         32.8         60.2         256           6274         2676940         2676243         660         pir.140339         Corynebacterium glutamicum (plutamicum 27.6         60.4         225           6275         2676940         2677377         438         sp:TIR2_YEAST         Saccharomyces crewisiae         24.2         62.1         124           6276         2677193         2676918         276         pir.C69281         Archaeoglobus fulgidus AF0251         50.0         86.0         50           6277         2679588         2677478         2121         gp:AF112535_3         Corynebacterium glutamicum 399.9         100.0         707 | SEQ   Initial   Terminal   ORF   db Match   Homologous gene   (%)   (% | SEQ Initial (a.a.)         Terminal (bp)         CRF (bp)         db Match (bp)         Homologous gene (bp)         Identity (bp)         Matched (bp)           8271 2674339 (a.a.)         2673338 (bp)         1002 gp:AF112536_1         Corynebacterium glutamicum gg.7 (bp)         99.7 (bp)         334           6272 2674804 2675289 (bp)         2675289 486 sp:FTNA_ECOLI (bp)         Escherichia coli K12 ftnA (bp)         31.5 (bp)         60.2 (bp)         256           6273 2676902 2676243 (bp)         2676249 750 gp:SCA32WHIH_A (bp)         Streptomyces coelicolor A3(2) 32.8 (bp)         32.8 (bp)         256           6276 2676940 2676243 (bp)         2676243 (bp)         2676243 (bp)         2676344 (bp)         27.6 (bp)         27.6 (bp)         27.6 (bp)           6276 2676940 2677377 (bp)         27.6 (bp)         27.7 (bp)         27.7 (bp)         27.7 (bp)         27.7 (bp)         27.7 (bp)           6276 277 2676948 276 (bp)         27.7 (bp) | SEQ         Initial         Terminal         ORF         db Match         Homologous gene         Identity (%)         Similarity (%)         Matched (%)           NOD (11)         (I1)         (ID)         < | Table 1 (continued)   SEQ   Initial   Terminal   CRF   Ab Match   Homologous gene   (%) | Table 1 (continued)   Terminal   ChPF   Ab Match   Homologous gene   Homologous gene   (%)   ( | Table 1 (continued)   SEQ   Initial   Terminal   ORF   db Match   Homologous gene   (%) | SEQ         Initial         Teminal         ORF         db Match         Homologous gene         (%)         Matched           (ND)         (Int)         (Int) | SEQ                           | SEG     | SEG                      |

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

	Function	hypothetical membrane protein	hypothetical membrane protein	hypothetical protein	transposase (IS1676)	major secreted protein PS1 protein precursor				transposase (IS1676)		proton/sodium-glutamate symport protein		ABC transporter		ABC transporter ATP-binding protein	hypothetical protein	hypothetical protein		oxidoreductase or dehydrogenase
	Matched length (a.a.)	84	122	254	496	355				900		438		873		218	84	42		196
	Similarity (%)	64.3	61.5	79.1	48.6	49.6				46.6		66.2		0.69		79.8	67.0	75.0		54.1
	Identity (%)	41.7	25.4	51.2	24.2	24.8				24.6		30.8		33.0		45.4	60.0	71.0		28.1
Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv3069	Helicobacter pylori J99 jhp1146	Bacillus subtilis 168 ycsl	Rhodococcus erythropolis	Corynebacterium glutamicum (Brevibacterium flavum) ATCC 17965 csp1				Rhodococcus erythropolis		Bacillus subtilis 168		Streptomyces coelicolor A3(2) SCE25.30	-	Staphylococcus aureus	Chlamydophila pneumoniae AR39 CP0987	Chlamydia muridarum Nigg TC0129		Streptomyces collinus Tu 1892 ansG
	db Match	pir:F70650	pir:D71843	sp:YCSI_BACSU	gp:AF126281_1	sp:CSP1_CORGL				gp:AF126281_1		sp:GLTT_BACCA		gp:SCE25_30		gp:SAU18641_2	PIR:F81516	PIR:F81737		prf:2509388L
	ORF (bp)	288	324	792	1365	1620	354	165	447	1401	768	1338	693	2541	891	802	273	141	678	672
	Terminal (nt)	2690437	2690760	2691564	2693053	2694918	2695279	2695718	2695320	2697212	2697383	2698194	2701612	2699926	2703356	2702487	2704586	2704975	2710555	2711308
	Initial (nt)	2690150	2690437	2690773	2691689	2693299	2694926	6296 2695554	2695766	2695812	6299 2698150	6300 2699531	2700920	2702466	2702466	2703194	2704314	2704835	2709878	2710637
	SEQ NO. (a.a.)	6290	6291	6292	6293	6294	6295		6297	6298	6239		6301	6302	6303	6304	6305	9089	6307	6308
	SEQ NO. (DNA)	2790	2791	2792	2793	2794	2795	2796	2797	2798	2799	2800	2801	2802	2803	2804	2805	2806	2807	2808

succinyl-CoA coenzyme A transferase

501

77.8

47.9

Clostridium kluyveri cat1 cat1

1539 sp.CAT1\_CLOKL

6326 2730916 2729378

transcriptional regulator

321

68.5

38.6

Azospirillum brasilense ATCC 29145 ntrC

1143 Sp.NIR3\_AZOBR

2732518

2731376

2827

											-					_	$\overline{}$		_	_
5	a.	_					mine 1- se		lor			ase		ase alpha		ase beta chain		duct		me A
10		Function	methyltransferase	hypothetical protein	hypothetical protein		UDP-N-acetylglucosamine carboxyvinyltransferase	hypothetical protein	transcriptional regulator		cysteine synthase	O-acetylserine synthase	hypothetical protein	succinyl-CoA synthetase alpha chain	hypothetical protein	succinyl-CoA synthetase beta chain		frenolicin gene E product		Succinvi-CoA coenzyme
15		Matched length (a.a.)	205	84	42		417	190	281		305	172	83	291	75	400		213		
20		Similarity (%)	51.2	66.0	75.0		75.3	84.2	0.69		84.6	79.7	65.1	79.4	43.0	73.0		71.8		
		Identity (%)	25.9	61.0	71.0		44.8	66.3	45.9		57.1	61.1	36.1	52.9	42.0	39.8		38.5		
25	(panu	ine.	losis	0	Nigg		sticus	losis	or A3(2)		sK	cysE2	ans R1	Mile Ph I	4PE1069	ည		vus frnE		
30	Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv0089	Chlamydia pneumoniae	Chlamydia muridarum Nigg TC0129		Acinetobacter calcoaceticus NCIB 8250 murA	Mycobacterium tuberculosis H37Rv Rv1314c	Streptomyces coelicolor A3(2) SC2G5.15c		Bacillus subtilis 168 cysK	Azotobacter vinelandii cysE2	Deinococcus radiodurans R1 DR1844	Coxiella burnetii Nine Mile Ph I sucD	Aeropyrum pernix K1 APE1069	Bacillus subtilis 168 sucC		Streptomyces roseofulvus frnE		
<i>35</i> <i>40</i>		db Match	Sp. Y089_MYCTU N	GSP:Y35814 C	PIR:F81737 T		sp:MURA_ACICA	Sp:Y02Y_MYCTU	gp:SC2G5_15		Sp.CYSK_BACSU E		gp:AE002024_10	s onco_coxed	PIR:F72706 A	sp:SUCC_BACSU E		gp:AF058302_5 s		
		ORF (bp)	525	273 (	141	195	1254	570	843	408	924	546	288	882	225	1194	360	735	819	
45		Terminal (nt)	2712374	2713453	2713842	2717993	2718436	2720319	2720385	2721295	2722857	2723609	2723770	2724478	2725843	2725384	2726786	2727399	2728207	
50		Initial (nt)	2711850	2713181	2713702	2718187	2719689	2719750	2721227	2721702	2721934	2723064	2724057	2725359	2725619	2726577	2727145	2728133	2729025	
		SEQ NO.	<del></del>	6310	6311	6312	6313	6314	6315	6316		6318		6320	6321		6323	6324	6325	
55		SEQ NO.	2809	2810	2811	2812	2813	2814	2815	2816	2817	2818	2819	2820	2821	2822	2823	2824	2825	

5'-phosphoribosyl-5-aminoimidazole synthetase

347

94.2

81.0

Corynebacterium ammoniagenes ATCC 6872 purM

gp:AB003158\_5

1074

2744881

2745954

6341

amidophosphoribosyl transferase

482

89.0

70.3

Corynebacterium ammoniagenes ATCC 6872 purF

gp:AB003158\_4

1482

2746083

2747564

6342

2842

1		$\overline{}$												
5	noi		system	ransport	sport system	sport system	orotein S-3			-		no acid		
10	Function		phosphate transport system regulatory protein	phosphate-specific transport component	phosphate ABC transport system permease protein	phosphate ABC transport system permease protein	phosphate-binding protein S-3 precursor	acetyltransferase		hypothetical protein	hypothetical protein	branched-chain amino acid aminotransferase	hypothetical protein	hypothetical protein
15	Matched length (a.a.)		213	255	292	325	369	315		344	225	529	352	58
20	Similarity (%)		81.7	82.8	82.2	78.5	26.0	0.09		55.2	74.2	56.0	79.0	81.0
	Identity (%)		46.5	58.8	51.4	50.2	40.0	34.3		24.7	44.9	28.6	58.5	58.6
<i>25</i> (pənu	J. G.		losis 2	osa pstB	ılosis	losis	losis	ır A3(2)		ırU	losis	CAT2	6872	losis
% See See See See See See See See See Se	Homologous gene		Mycobacterium tuberculosis H37Rv Rv0821c phoY-2	Pseudomonas aeruginosa pstB	Mycobacterium tuberculosis H37Rv Rv0830 pstA1	Mycobacterium tuberculosis H37Rv Rv0829 pstC2	Mycobacterium tuberculosis H37Rv phoS2	Streptomyces coelicolor A3(2) SCD84.18c		Bacillus subtilis 168 bmrU	Mycobacterium tuberculosis H37Rv Rv0813c	Solanum tuberosum BCAT2	Corynebacterium ammoniagenes ATCC 6872 ORF4	Mycobacterium tuberculosis H37Rv Rv0810c
40	db Match		pir.E70810	pir.S68595	gp:MTPSTA1_1	pir.A70584	pir.H70583	gp:SCD84_18		sp:BMRU_BACSU	pir.E70809	gp:AF193846_1	gp:AB003158_6	pir.B70809
	ORF (bp)	807	732	897	921	1014	1125	876	783	1095	687	942	1101	213
45	Terminal (nt)	2731424	2733367	2733455	2734264	2735202	2736414	2737836	2739553	2739556	2741356	2741636	2743785	2744222
50	Initial (nt)	2732230	2732636	2734351	2735184	2736215	2737538	2738711	2738771	2740650	2740670	2742577	2742685	2744010
	SEQ NO.	6328	6329	6330	6331	6332	6333	6334	6335	6336	6337	6338	6339	6340
55	SEO NO.	2828	2829	2830	2831	2832	2833	2834	2835	2836	2837	2838	2839	2840

5	Ę		, vic	ne protein		synthetase		synthetase			se	ov.			nsporter	idase
10	Function	hypothetical protein	hypothetical protein	hypothetical membrane protein	hypothetical protein	5'-phosphoribosyl-N- formylglycinamidine synthetase		5-phosphoribosyl-N- formylglycinamidine synthetase	hypothetical protein		gluthatione peroxidase	extracellular nuclease		hypothetical protein	C4-dicarboxylate transporter	dipeptidyl aminopeptidase
15	Matched length (a.a.)	124	315	217	42	763		223	79		158	965		211	414	269
20	Similarity (%)	75.8	94.0	87.1	71.0	89.5		93.3	93.7		77.9	51.5		68.7	81.6	70.6
	Identity (%)	57.3	75.9	67.7	64.0	77.6		80.3	81.0		46.2	28.0		37.4	49.0	41.8
<sub>25</sub>		.s	2	72		72		72	72			P636		sis	T2	dapb1
© Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv0807	Corynebacterium ammoniagenes ATCC 6872 ORF2	Corynebacterium ammoniagenes ATCC 6872 ORF1	Sulfolobus solfataricus	Corynebacterium ammoniagenes ATCC 6872 purL		Corynebacterium ammoniagenes ATCC 6872 purQ	Corynebacterium ammoniagenes ATCC 6872 purorf		Lactococcus lactis gpo	Aeromonas hydrophila JMP636 nucH		Mycobacterium tuberculosis H37Rv Rv0784	Salmonella typhimurium LT2 dctA	Pseudomonas sp. WO24 dapb1
35		Myc. H37	Coryne ammol	Coryno ammo ORF1		Cory amm purL		Coryn ammc purd	Coryne ammor purorf		Lact	Aeron		Myc H37	Salm	Pse
40	db Match	pir:H70536	gp:AB003158_2	gp:AB003158_1	GP:SSU18930_21 4	gp:AB003162_3		gp:AB003162_2	gp:AB003162_1		prf:2420329A	prf.2216389A		pir:C70709	sp:DCTA_SALTY	prf:2408266A
	ORF (bp)	375	1017	741	186	2286	720	699	243	522	477	2748	276	687	1338	2118
45	Terminal (nt)	2747683	2749111	2749162	2752103	2750027	2753121	2752327	2752995	2753819	2753328	2756739	2757126	2757129	2757863	2759532
50	Initial (nt)	2748057	2748095	2749902	2751918	2752312	2752402		2753237	2753298	2753804	2753992	2756851	2757815	2759200	2761649
	SEQ NO.	6343	6344	6345	6346	6347	6348	-	6350	6351	6352	6353	6354	6355	6356	6357
55	SEQ NO.	2843	2844	2845	2846	2847	2848	2849	2850	2851	2852	2853	2854	2855	2856	2857

5	Function		5-phosphoribosyl-4-N- succinocarboxamide-5-amino imidazole synthetase	adenylosuccino lyase	aspartate aminotransferase	5'-phosphoribosylglycinamide synthetase	histidine triad (HIT) family protein		hypothetical protein	di-/tripeptide transpoter	adenosylmethionine-8-amino-7- oxononanoate aminotransferase or 7,8-diaminopelargonic acid aminotransferase	dethiobiotin synthetase	two-component system sensor histidine kinase	two-component system regulatory protein	transcriptional activator	metal-activated pyridoxal enzyme or low specificity D-Thr aldolase
	Matched length (a.a.)		294	477	395	425	136		243	469	423	224	335	231	249	382
20	Similarity (%)		89.1	95.0	62.3	86.4	80.2		56.4	67.6	98.8	9.66	70.5	72.7	69.5	53.9
	Identity (%)		70.1	85.3	28.1	71.1	23.7		26.8	30.1	95.7	98.7	31.3	42.0	37.4	30.9
25 (juned)	ene		: 6872	: 6872	s ATCC	: 6872	s u296a		eri orf3	bsp. lactis	amicum m) MJ233	amicum п) MJ233	1plasmid	drrA	tipA:	83
& Table 1 (continued)	Homologous gene		Corynebacterium ammoniagenes ATCC 6872 purC	Corynebacterium ammoniagenes ATCC 6872 purB	Sulfolobus solfataricus ATCC 49255	Corynebacterium ammoniagenes ATCC 6872 purD	Mycobacterium leprae u296a		Methanosarcina barkeri orf3	Lactococcus lactis subsp. lactis	Corynebacterium glutamicum (Brevibacterium flavum) MJ233 bioA	Corynebacterium glutamicum (Brevibacterium flavum) MJ233 bioD	Lactococcus lactis M71plasmid pND306	Thermotoga maritima drrA	Streptomyces lividans tipA	Arthrobacter sp. DK-38
35									2					Τ		∢
40	db Match		gp:AB003161_3	gp:AB003161_2	sp:AAT_SULSO	gp:AB003161_1	Sp:YHIT_MYCLE		pir.S62195	sp:DTPT_LACLA	sp:BIOA_CORGL	sp:BIOD_CORGL	gp:AF049873_3	prf:2222216A	Sp:TIPA_STRLI	prf.2419350A
	ORF (bp)	624	891	1428	1158	1263	414	435	753	1356	1269	672	1455	705	753	1140
45	Terminal (nt)	2761829	2761785	2763504	2764978	2766158	2767993	2767703	2768343	2769156	2771982	2772660	2772644	2774110	2774937	2775740
50	Initial (nt)	2762452	2762675	2764931	2766135	2767420	2767580	2768137	2769095	2770511	2770714	2771989	2774098	2774814	2775689	2776879
	SEQ NO. (a.a.)	6358	6329	6360	6361	6362	6363	6364	6365	6366	6367	6368	6369	6370	6371	6372
55	SEQ NO.	2858	2859	2860	2861	2862	2863	2864	2865	2866	2867	2868	2869	2870	2871	2872

. 5		Function	pyruvate oxidase	multidrug efflux protein	transcriptional regulator	hypothetical membrane protein		3-ketosteroid dehydrogenase	transcriptional regulator, LysR family	hypothetical protein	hypothetical protein		hypothetical protein	hypothetical membrane protein	transcription initiation factor sigma	trehalose-6-phosphate synthase		trehalose-phosphatase	glucose-resistance amylase regulator	high-affinity zinc uptake system protein
15		) is ed													7					
		Matched length (a.a.)	574	504	92	421		303	232	278	288		4	464	155	487		245	344	353
20		Similarity (%)	75.8	68.9	68.5	78.4		62.1	0.69	52.9	55.6		50.7	64.0	50.3	66.7		57.6	60.2	46.7
		Identity (%)	46.3	33.3	30.4	45.6		34.3	37.1	28.4	26.7		28.6	36.0	32.3	38.8		27.4	24.7	22.4
<i>25</i>	(nan)	ē	хВ	plasmid	ည	osis		lis SQ1	4	losis	A		idney	losis	rdB	oombe		зВ	ρA	e Rd
	lable I (conlinued)	Homologous gene	Escherichia coli K12 poxB	Staphylococcus aureus plasmid pSK23 qacB	Escherichia coli K12 ycdC	Mycobacterium tuberculosis H37Rv Rv2508c		Rhodococcus erythropolis SQ1 kstD1	Bacillus subtilis 168 alsR	Mycobacterium tuberculosis H37Rv Rv3298c lpqC	Bacillus subtilis 168 ykrA		Oryctolagus cuniculus kidney cortex rBAT	Mycobacterium tuberculosis H37Rv Rv3737	Streptomyces griseus hrdB	Schizosaccharomyces pombe tps1		Escherichia coli K12 otsB	Bacillus megaterium ccpA	Haemophilus influenzae Rd Hi0119 znuA
35		,		S	-	ΣÏ			_	ZI	<u>a</u>		0 6	21	S					
40		db Match	gp:ECOPOXB8G	prf.2212334B	sp:YCDC_ECOLI	pir.D70551		gp:AF096929_2	SP:ALSR_BACSU	pir.C70982	pir.C69862		pir.A45264	pir.870798	pir.S41307	sp:TPS1_SCHPO		sp.OTSB_ECOLI	sp:CCPA_BACME	sp:ZNUA_HAEIN
		ORF (bp)	1737	1482	531	1320	2142	096	705	813	813	459	399	1503	327	1455	513	768	1074	942
45		Terminal (nt)	2776768	2780446	2780969	2782315	2782340	2784656	2785651	2788594	2788587	2789477	2790550	2792448	2792857	2794327	2794812	2795637	2795676	2797806
50		Initial (nt)	2778504	2778965	2780439	2780996	2784481	2785615	2786355	2787782	2789399	2789935	2790152	2790946	2792531	2792873	2794300	2794870	2796749	2796865
		SEQ NO.	6373	6374	6375	6376	6377	6378	6379	6380	6381	6382	6383	6384	6385	6386	6387	6388	6389	6390
55		SEQ NO.	2873	2874	2875	2876	2877	2878	2879	2880	2881	2882	2883	2884	2885	2886	2887	2888	2889	2890

5
10
15
20
25
30
35
40
45
50

Function	ABC transporter	hypothetical membrane protein	transposase (ISA0963-5)		3-ketosteroid dehydrogenase		lipopolysaccharide biosynthesis protein or oxidoreductase or dehydrogenase	dehydrogenase or myo-inositol 2- dehydrogenase	shikimate transport protein	shikimate transport protein	transcriptional regulator	ribosomal RNA ribose methytase or tRNA/rRNA methyltransferase	cysteinyl-tRNA synthetase	PTS system, enzyme II sucrose protein (sucrose-specific IIABC component)	sucrose 6-phosphate hydrolase or sucrase	glucosamine-6-phosphate isomerase	N-acetylglucosamine-6-phosphate deacetylase
Matched length (a.a.)	223	135	303		561		204	128	282	130	212	334	464	668	473	248	368
Identity Similarity (%)	63.2	87.4	52.5		62.0		56.4	69.5	67.5	80.8	55.7	47.3	68.8	77.0	56.9	69.4	60.3
Identity (%)	31.4	0.09	23.4		32.1		34.3	35.2	30.5	43.1	32.6	22.8	42.2	47.0	35.3	38.3	30.2
Homalogous gene	Staphylococcus aureus 8325-4 mreA	Mycobacterium tuberculosis H37Rv Rv2060	Archaeoglobus fulgidus		Rhodococcus erythropolis SQ1 kstD1		Thermotoga maritima MSB8 bpIA	Bacillus subtilis 168 idh or iolG	Escherichia coli K12 shiA	Escherichia coli K12 shiA	Streptomyces coelicolor A3(2) SC5A7.19c	Saccharomyces cerevisiae YOR201C PET56	Escherichia coli K12 cysS	Lactococcus lactis sacB	Clostridium acetobutylicum ATCC 824 scrB	Escherichia coli K12 nagB	Vibrio furnissii SR1514 manD
db Match	gp:AF121672_2	pir:E70507	pir:A69426	·	gp:AF096929_2		pir.872359	sp:MI2D_BACSU	SP:SHIA_ECOLI	Sp.SHIA_ECOLI	gp:SC5A7_19	sp:PT56_YEAST	sp:SYC_ECOLI	prf.2511335C	gp:AF205034_4	sp:NAGB_ECOL!	sp:NAGA_VIBFU
ORF (bp)	9	555	1500	201	1689	747	618	435	855	426	654	939	1380	1983	1299	759	1152
Terminal (nt)	2798509	2799391	2801034	2801313	2801558	2803250	2804074	2804676	2805113	2806016	2806599	2807426	2808399	2809824	2811960	2813279	2814081
Initial (nt)	2797820	2798837	2799535	2801113	2803246	2803996	2804691	2805110	2805967	2806441	2807252	2808364	2809778	2811806	2813258	2814037	2815232
SEQ NO (a.a.)	6391	6392	6393	6394	6395	9689	2689	6398	6333	6400	6401	6402	6403	6404	6405	6406	6407
SEQ NO. (DNA)	2891	2892	2893	2894	2895	2896	2897	2898	2899	2900	2901	2902	2903	2904	2905	2906	2907

10	Function	dihydrodipicolinate synthase	glucokinase	N-acetylmannosamine-6-phosphate epimerase	A more and a more and	sialidase precursor	L-asparagine permease operon repressor	dipeptide transporter protein or heme-binding protein	dipeptide transport system permease protein	oligopeptide transport ATP-binding protein	oligopeptide transport ATP-binding protein	homoserine/homoserin lactone efflux protein or lysE type translocator	leucine-responsive regulatory protein		hypothetical protein	hypothetical protein	transcription factor
15	Matched length (a.a.)	298 c	321	220		439	222 L	560	342	314	258	193	142		152	235	157 t
20	Similarity (%)	62.1	57.6	68.6		50.3	57.2	51.4	64.3	78.3	78.7	62.7	66.2		86.2	71.5	91.1
	Identity (%)	28.2	28.7	36.4		24.8	26.6	22.5	31.9	46.5	43.4	28.5	31.0		55.9	46.4	73.3
25 Table 1 (continued)	ыз депе	12 dapA	icolor A3(2)	gens NCTC		iridifaciens	R	4 dppA	4 dappB	8 орр	оррF	12 rhtB	ponicum Irp		oerculosis	oerculosis	oerculosis
·	Homologous gene	Escherichia coli K12 dapA	Streptomyces coelicolor A3(2) SC6E10.20c glk	Clostridium perfringens NCTC 8798 nanE		Micromonospora viridifaciens ATCC 31146 nadA	Rhizobium etli ansR	Bacillus firmus OF4 dppA	Bacillus firmus OF4 dappB	Bacillus subtilis 168 oppD	Lactococcus lactis oppF	Escherichia coli K12 rhtB	Bradyrhizobium japonicum Irp		Mycobacterium tuberculosis H37Rv Rv3581c	Mycobacterium tuberculosis H37Rv Rv3582c	Mycobacterium tuberculosis H37Rv Rv3583c
40	db Match	p.DAPA_ECOLI E	sp.GLK_STRCO S	rf.2516292A 8		p:NANH_MICV! A	p:AF181498_1 R	p:BFU64514_1 B	p:DPPB_BACFI B	p:OPPD_BACSU B	P:OPPF_LACLA	p:RHTB_ECOL! E	rf.2309303A B		pir.C70607	p:Y18T_MYCTU N	pir:H70803
	ORF (bp)	936 sp	ds 606	696 prf	177	1215 sp	729 gp	1608 gp	951 sp	1068 sp	816 sp	621 sp	483 prf	360	480 pir	768 sp	594 pir
45	Terminal (nt)	2816393	2817317	2818058	2818137	2818350	2819557	2822191	2823337	2825341	2826156	2826215	2827404	2827458	2827904	2828379	2829156
50	Initial (nt)	2815458	2816409	2817363	2818313	2819564	2820285	2820584	2822387	2824274	2825341	2826835	2826922	2827817	2828383	2829146	2829749
	SEQ NO. (a.a.)	6408	6409	6410	6411	6412	6413	6414	6415	6416	6417	6418	6419	6420	6421	6422	6423
55	SEQ NO. (DNA)	2908	2909	2910	2911	2912	2913	2914	2915	2916	2917	2918	2919	2920	2921	2922	2923

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

_			<del></del>																	
	Function	two-component system response regulator	two-component system sensor histidine kinase		DNA repair protein RadA	hypothetical protein	hypothetical protein	p-hydroxybenzaldehyde dehydrogenase		mitochondrial carbonate dehydratase beta	A/G-specific adenine glycosylase			L-2.3-butanediol dehydrogenase				hypothetical protein	virulence factor	virulence factor
	Matched length (a.a.)	223	341		463	345	231	471		210	283			258				97	66	72
	Similarity (%)	70.0	67.7		74.3	73.3	53.3	85.1		.00	7.07			9.66		,		69.1	63.0	55.0
	Identity (%)	43.5	29.3		41.5	40.3	29.4	59.5	×	36.7	48.4			99.2				48.5	57.0	54.0
Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv3246c mtrA	Escherichia coli K12 baeS		Escherichia coli K12 radA	Bacillus subtilis 168 yacK	Mycobacterium tuberculosis H37Rv Rv3587c	Pseudomonas putida NCIMB 9866 plasmid pRA4000		Chlamydomonas reinhardtii ca1	Streptomyces antibioticus IMRU 3720 mutY			Brevibacterium saccharolyticum				Mycobacterium tuberculosis H37Rv Rv3592	Pseudomonas aeruginosa ORF24222	Pseudomonas aeruginosa ORF25110
	db Match	prf.2214304A	sp:BAES_ECOLI		sp:RADA_ECOLI	sp:YACK_BACSU	pir.D70804	gp:PPU96338_1		pir:T08204	gp:AF121797_1			gp:AB009078_1				pir:E70552	GSP:Y29188	GSP:Y29193
	ORF (bp)	723	1116	582	1392	1098	289	1452	147	621	879	1155	306	774	324	741	312	291	420	213
	Terminal (nt)	2830779	2831894	2832666	2834181	2835285	2835283	2836048	2837591	2837956	2839521	2840716	2840758	2841848	2842453	2843233	2843716	2843432	2845558	2846101
	initial (nt)	2830057	2830779	2832085	2832790	2834188	2835969	2837499	2837737	2838576	2838643	2839562	2841063	2841075	2842130	2842493	2843405	2843722	2845139	2845889
	SEQ NO. (a.a.)	6424	6425	6426	6427	6428	6429	6430	6431	6432	6433	6434	6435	6436	6437	6438	6439	6440	6441	6442
	SEQ NO. (DNA)	2924	2925	2926	2927	2928	2929	2930	2931	2932	2933	2934	2935	2936	2937	2938	2939	2940	2941	2942

hypothetical membrane protein

138

φ. 69

29.0

Mycobacterium leprae MLCB2548.04c

gp:MLCB2548\_4

465 798

2864384

2864848

6458

2958

2864421 2863624

6457

2957

2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase

158

69.0

42.4

Methylobacterium extorquens AM1 folK

SP.HPPK\_METEX

477

6459 2865343 2864867

2959

dihydroneopterin aldolase dihydropteroate synthase

118 268

69.5 75.0

38.1

51.5

Mycobacterium leprae folP Bacillus subtilis 168 folB

sp:FOLB\_BACSU

390 837

6460 2865735 2865346

2960 2961

gp: AB028656\_1

2866567 2865731

6461

г									$\neg \neg$	$\neg$					
	د		osphatase / se	ate		nase					protein		8	ie ligase	
	Functio	virulence factor	ClpC adenosine triph ATP-binding proteina	inosine monophosphi dehydrogenase	transcription factor	phenol 2-manooxyge					lincomycin resistance	hypothetical protein	lysyl-tRNA synthetas	pantoatebeta-alanir	
	Matched length (a.a.)	55	832	469	316	680					481	240	511	268	
	Similarity (%)	75.0	86.2	70.2	62.7	6.09					100.0	55.8	71.2	52.6	
	Identity (%)	74.0	58.5	37.1	24.7	33.5		ļ			100.0	26.7	41.7	29.9	
ਰੂ				_	s nitR	202					Ë	sis	s lysS	mn:	
Table 1 (continue	Homologous gene	Pseudomonas aeruginosa ORF25110	Bacillus subtilis 168 mecB	Bacillus cereus ts-4 impdh	Rhodococcus rhodochrous	Trichosporon cutaneum A 46490		-			Corynebacterium glutamic ImrB	Mycobacterium tuberculos H37Rv Rv3517	Bacillus stearothermophilu	Corynebacterium glutamic ATCC 13032 panC	
	db Match	GSP:Y29193	sp:MECB_BACSU	gp:AB035643_1		rRICU					gp:AF237667_1	pir:G70807	gp:AB012100_1		
	ORF (bp)	321	2775	1431	1011	1785	1716	1941	1722	162	1443	951	1578	798	693
	Terminal (nt)	2846506	2844166	2848659	2849779	2851815	2853732	2855709	2857516	2859205	2857613	2859195	2860505	2862132	2862929
	Initial (nt)	2846186	2846940	2847229	2848769	2850031	2852017	2853769	2855795	2859044	2859055	2860145	2862082	2862929	2863621
	SEQ NO.	6443	6444	6445	6446	6447	6448	6448	6450	6451	6452		6454	6455	6456
	SEQ NO.	2943	2944	2945	2946	2947	2948	2949	2950	2951	2952	2953	2954	2955	2956
	Table 1 (continued)	SEQ Initial Terminal ORF db Match Homologous gene (%) (mt) (ht) (bp)	SEQ Initial (a.a.)         Terminal (a.a.)         ORF (bp)         db Match         Homologous gene (w)         Identity (w)         Similarity length (a.a.)         Matched (a.a.)           NO. (a.a.)         (a.b.)         (bp)         (bp)         Pseudomonas aeruginosa         74.0         75.0         55         virulence factor	SEQ Initial (a.a.)         Terminal (bp)         ORF (bp)         db Match         Homologous gene (%)         Identity (%)         Similarity length (a.a.)         Matched (a.a.)           6443         2846506         321         GSP:Y29193         Pseudomonas aeruginosa ORF25110         74.0         75.0         55         virulence factt CIPC adenosii           6444         2846940         284166         2775         sp.MECB_BACSU         Bacillus subtilis 168 mecB         58.5         86.2         832         ATP-binding F	SEQ (nt)         Initial (nt)         Terminal (nt)         ORF (bp)         db Match         Homologous gene (%)         Identity (%)         Similarity length (a.a.)         Matched (a.a.)           6443         2846186         2846506         321         GSP:Y29193         Pseudomonas aeruginosa ORF25110         74.0         75.0         55           6444         2846940         2844166         2775         sp:MECB_BACSU         Bacillus subtilis 168 mecB         58.5         86.2         832           6445         2847229         2848659         1431         gp:AB035643_1         Bacillus cereus ts-4 impdh         37.1         70.2         469	SEQ (nt) (nt) (nt) (nt) (nt) (nt) (nt) (nt)	SEQ NO. (nt)         Initial (nt)         Terminal (nt)         ORF (pp)         db Match         Homologous gene (%)         Identity (%)         Similarity (%)         Matched (%)         Matched (%)<	SEQ         Initial (nt)         Terminal ORF         db Match         Homologous gene         Identity (%)         Similarity (%)         Matched (%)           NO. (nt)         (nt)         (nt)         (bp)         db Match         Homologous gene         14.0         75.0         55           6443         2846186         2846506         321         GSP:Y29193         Pseudomonas aeruginosa         74.0         75.0         55           6444         2846940         2844166         2775         sp:MECB_BACSU         Bacillus subtilis 168 mecB         58.5         86.2         832           6445         2848694         1431         gp:AB035643_1         Bacillus cereus ts-4 impdh         37.1         70.2         469           6446         2848769         2849779         1011         pir.JC6117         Rhodococcus rhodochrous nitR         24.7         62.7         316           6446         2850031         2851815         1785         sp:PH2M_TRICU         46490         60.9         60.9         60.9         60.9         60.9         60.9         60.9         60.9         60.9         60.9         60.9         60.9         60.9         60.9         60.9         60.9         60.9         60.9         60.9	SEQ         Initial (nt)         Terminal (nt)         ORF (nt)         db Match (bp)         Homologous gene (%)         Identity (%)         Similarity (%)         Matched (%)           6443         2846186         2846506         321         GSP:Y29193         Pseudomonas aeruginosa (%)         75.0         55           6444         2846186         2844166         2775         sp:MECB_BACSU         Bacillus subtilis 168 mecB (%)         75.0         55           6445         2846549         2844166         1431         gp:AB035643_1         Bacillus cereus ts-4 impdh (%)         37.1         70.2         469           6446         2848729         2849779         1011         pir.JC6117         Rhodococcus rhodochrous nitR (24.7)         62.7         316           6446         2852017         2853732         1716         Trichosporon cutaneum ATCC (33.5         60.9         680           6449         2853769         2855709         1941         178         178         178         178	Table 1 (continued)   Terminal   ORF   db Match   Homologous gene   (%)   (%	SEQ Initial (a.a.)         Terminal (bp)         GB Match (bp)         Abmatch (bp)         Homologous gene (%)         Identity (%)         Similarity (Bp) (ha)         Matched (%)         Matched (%)	SEQ         Initial         Terminal (nt)         ORF (pp)         db Match (pp)         Homologous gene         Identify (%)         Similarity (%)         Matched (%a)           NOO (nt)         (nt)         (nt)         (nt)         (nt)         (nt)         (nt)         (pp)         (pp	SEQ         Initial         Terminal (nt)         ORF         db Match         Homologous gene         Identity (%)         Similarity (%)         Matched (%)           NOD (1xt)         (xt)         (xt)         (bp)         db Match         Homologous gene         Identity (%)         (xt)         (xt)           6443         2846186         2846506         321         GSP:Y29193         Pseudomonas aeruginosa         74.0         75.0         55           6444         2846940         2844166         2775         sp:MECB_BACSU         Bacillus subtilis 168 mecB         58.5         86.2         832           6446         2846940         2844166         2775         sp:MECB_BACSU         Bacillus subtilis 168 mecB         58.5         86.2         469           6446         2848769         1431         gp:AB035643_1         Bacillus subtilis 168 mecB         58.5         86.2         469           6446         2848769         2849779         1011         pir.JC6117         Rhodococcus thodochrous nitR         24.7         62.7         316           6449         285709         1941         Trichosporton cutaneum ATCC         33.5         60.9         680           6450         2855795         1841         Trichosporton c	SEQ         Initial         Terminal         ORF         db Match         Homologous gene         Identity (%)         Iminarity (%)         Matched (%)           NO.         (nt)         (nt)         (pp)         db Match         Homologous gene         (%) <td< td=""><td>SEQ         Initial         Terminal (hp)         CRF         db Match         Homologous gene         Identity (%)         Similarity length (%)         Matched (%)           NO.         (nt)         (nt)         (nt)         (hp)         db Match         Homologous gene         (%)</td></td<>	SEQ         Initial         Terminal (hp)         CRF         db Match         Homologous gene         Identity (%)         Similarity length (%)         Matched (%)           NO.         (nt)         (nt)         (nt)         (hp)         db Match         Homologous gene         (%)

											<u> </u>		— Т		т					
5		Function	GTP cyclohydrolase I		cell division protein FtsH	hypoxanthine phosphoribosyltransferase	cell cycle protein MesJ or cytosine deaminase-related protein	D-alanyl-D-alanine carboxypeptidase	inorganic pyrophosphatase		spermidine synthase	hypothetical membrane protein	hypothetical protein	hypothetical protein	hypothetical protein	PTS system, beta-glucosides- permease II ABC component		ferredoxin reductase	hypothetical protein	bacterial regulatory protein, marR family
			GTP		<u>e</u>	pyd Spodd	dear	D-al carb	inor	]	sbei	hyp	hyp.	hyp	hyp	PTG		fer	qyd	fam
15		Matched length (a.a.)	188		782	165	310	459	159		507	132	144	173	202	68		411	6	135
20		Similarity (%)	86.2		69.0	83.0	66.8	51.4	73.6		80.7	86.4	63.2	60.1	72.3	59.6		9.69	73.2	59.3
		Identity (%)	9.09		56.0	51.5	41.0	27.2	49.7		56.0	38.6	36.8	36.4	44.6	30.3		38.0	46.4	26.7
25						960									*				2)	ORF
<i>30</i>	Table 1 (continued)	Homologous gene	Bacillus subtilis 168 mtrA			Salmonella typhimurium GP660 hprt	Mycobacterium tuberculosis H37Rv Rv3625c	Actinomadura sp. R39 dac	Escherichia coli K12 ppa		Mycobacterium tuberculosis H37Rv speE	Mycobacterium tuberculosis H37Rv Rv2600	Mycobacterium tuberculosis H37Rv Rv2599	Mycobacterium tuberculosis H37Rv Rv2598	Mycobacterium tuberculosis H37Rv Rv2597	Bacillus subtilis 168 bgIP		Nocardioides sp. KP7 phdD	Streptomyces coelicolor A3(2) SCH69.09c	Burkholderia pseudomallei ORF E
40		db Match	sp.GCH1_BACSU			gp:AF008931_1	sp:YZC5_MYCTU	sp:DAC_ACTSP	SP.IPYR ECOLI		pir:H70886	sp:Y0B1_MYCTU	sp:Y0B2_MYCTU	sp:Y0B3_MYCTU	sp:Y0B4_MYCTU	sp.PTBA_BACSU		gp:AB017795_2	g_63HCSCH69_9	prf.2516298U
		ORF (bp)	588	915	2580	582	891	1233	474	219	1539	399	411	498	609	249	264	1233	288	444
45		Terminal (nt)	2866586	2868385	2867169	2869863	2870499	2871445	2873399	2873393	2873905	2875434	2875870	2876280	2876777	2877455	2877595	2878478	2880252	2880987
50		Initial (nt)	2867173	2867471	2869748	2870444	2871389	2872677	2872926			2875832	2876280	2876777	2877385	2877703	2877858	2879710	2879965	2880544
		SEQ NO.	6462	6463	6464	6465	6466	6467	6468	6469	6470	6471	6472	6473	6474	6475	6476	6477	6478	6479
<i>55</i>		SEQ NO.	_	2963	2964	2965	2966	2967	2968	-	+ -	2971	2972	2973	2974	2975	2976	2977	2978	2979

5		Function	peptide synthase		phenylacetaldehyde dehydrogenase	hypothetical protein	hypothetical protein	hypothetical protein	heat shock protein or chaperon or groEL protein						
15		Matched length (a.a.)	1241 pept		488 pher	241 hypo	54 hyp	31 hypo	548 hear						
20			51.6		63.7	79.7	63.0	80.0	100.0						
		Identity Similarity (%)	28.4		35.0	57.3	62.0	74.0	99.5						
25	ntinued)	gene	sporus cpsB		padA	ıi Cj0604	rculosis	rculosis	ım MJ-233						
30	Table 1 (continued)	Homologous gene	Streptomyces roseosporus cpsB		Escherichia coli K12 padA	Campylobacter jejuni Cj0604	GP:MSGTCWPA_1 Mycobacterium tuberculosis	GP:MSGTCWPA_1 Mycobacterium tuberculosis	Brevibacterium flavum MJ-233						
35		atch					CWPA_1 N	CWPA_1 N							
40		db Match	prf.2413335A		prf.2310295A	gp:CJ11168X2_25	GP:MSG1	GP:MSG1	gsp:R94368						
		ORF (bp)	3885	1461	1563	918	162	177	1644	180	1209	963	1986	2454	
<b>45</b>		Terminal (nt)	2884882	2881844	2884935	2886916	2890346	2890553	2888897	2890751	2890930	2892138	2893100	2895072	
50		Initial (nt)	2880998	2883304	2886497	2887833	2890185	2890377	6486 2890540	2890930	2892138	2893100	2895085	6491 2897525	
		SEQ NO. (a.a.)	6480	6481	6482	6483	6484	6485		6487	6488	6489	6490	6491	
		0 0 8	100	=	22	33	7	35	မွ	37	88	60	8	7	1

Function	peptide synthase		phenylacetaldehyde dehydrogenase	hypothetical protein	hypothetical protein	hypothetical protein	heat shock protein or chaperon or groEL protein							hypothetical protein			peptidase			Na+/H+ antiporter or multiple resistance and pH regulation related protein A or NADH dehydrogenase
Matched length (a.a.)	1241		488	241	24	31	548							1236			447			767
Similarity (%)	51.6		63.7	79.7	63.0	80.0	100.0							42.3			68.0			68.3
Identity (%)	28.4		35.0	57.3	62.0	74.0	99.5							21.7			37.1			35.6
Homologous gene	Streptomyces roseosporus cpsB		Escherichia coli K12 padA	Campylobacter jejuni Cj0604	Mycobacterium tuberculosis	Mycobacterium tuberculosis	Brevibacterium flavum MJ-233							Homo sapiens MUC5B			Mycobacterium tuberculosis H37Rv Rv2522c			Staphylococcus aureus mnhA
db Match	prf:2413335A		prf.2310295A	gp:CJ11168X2_25	GP:MSGTCWPA_1	GP:MSGTCWPA 1	gsp:R94368							prf.2309326A			pir:G70870		*	prf:2504285B
ORF (bp)	3885	1461	1563	918	162	177	1644	180	1209	963	1986	2454	2799	3591	2775	612	1371	579	909	3057
Terminal (nt)	2884882	2881844	2884935	2886916	2890346	2890553	2888897	2890751	2890930	2892138	2893100	2895072	2897528	2900330	2903964	2906639	2908885	2909788	2909231	2913228
Initial (nt)	2880998	2883304	2886497	2887833	2890185	2890377		2890930	2892138	2893100	2895085	2897525	2900326	2903920	2906738	2907250	2907515	2909210	2909830	6499 2910172
SEQ NO.	6480	6481	6482	6483	6484	6485	6486	6487	6488	6489	6490	6491	6492	6493	6494	6495	6496	6497	6498	6499
SEQ NO.	2980	2981	2982	2983	2984	2985	2986	2987	2988	2989	2990	2991	2992	2993	2994	2995	2996	2997	2998	2999

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

	Function	Na+/H+ antiporter or multiple resistance and pH regulation related protein C or cation transport system protein	Na+/H+ antiporter or multiple resistance and pH regulation related protein D	Na+/H+ antiporter or multiple resistance and pH regulation related protein E	K+ efflux system or multiple resistance and pH regulation related protein F	Na+/H+ antiporter or multiple resistance and pH regulation related protein G	hypothetical protein	hypothetical protein		polypeptide deformylase	hypothetical protein	acetyltransferase (GNAT) family or N terminal acetylating enzyme			exodeoxyribonuclease III or exonuclease	cardiolipin synthase
	Matched length (a.a.)	104	523	161	11	121	178	334		184	71	339			31	513
	Similarity (%)	81.7	72.1	6.09	66.2	63.6	54.5	61.7		6.09	70.4	54.2			6.65	62.0
	Identity (%)	44.2	35.2	26.7	32.5	25.6	24.7	27.0		37.5	47.9	31.3			30.8	27.9
Table 1 (continued)	Homologous gene	Bacillus firmus OF4 mrpC	Bacillus firmus OF4 mrpD	Bacillus firmus OF4 mrpE	Rhizobium meliloti phaF	Staphylococcus aureus mnhG	Mycobacterium tuberculosis H37Rv ipV	Escherichia coli K12 ybdK		Bacillus subtilis 168 def	Mycobacterium tuberculosis H37Rv Rv0430	Mycobacterium tuberculosis H37Rv Rv0428c			Salmonella typhimurium LT2 xthA	Bacillus firmus OF4 cls
i	db Match	gp:AF097740_3	gp:AF097740_4	gp.AF097740_5	prf.2416476G	prf.2504285H	pir.D70594	sp:YBDK_ECOLI		sp:DEF_BACSU	pir.D70631	pir:870631			gp:AF108767_1	gp:BFU88888_2
	ORF (bp)	489	1668	441	273	378	594	1128	663	579	252	1005	699	630	789	1500
	Terminal (nt)	2913723	2915416	2915922	2916201	2916582	2917024	2917630	2918819	2920293	2919490	2921290	2919808	2920220	2922108	2923617
	Initial (nt)	2913235	2913749	2915482	2915929	2916205	2917617	2918757	2919481	2919715	2919741	2920286	2920476	2920849	2921320	2922118
	SEQ NO (a.a.)	6500	6501	6502	6503	6504	6505	6506	6507	6508	6099	6510	6511	6512	6513	6514
	SEQ NO. (DNA)	3000	3001	3002	3003	3004	3005	3006	3007	3008	3009	3010	3011	3012	3013	3014

						_				
SEQ NO. (DNA)	SEQ NO.	Initial (nt)	Terminal (nt)	ORF (bp)	db Match	Homologous gene	Identity (%)	Similarity (%)	Matched length (a.a.)	Function
3015	6515	2924191	2924844	654		-				
3016			2923954	1194	sp:BCR_ECOLI	Escherichia coli K12 bcr	31.6	67.2	393	membrane transport protein or bicyclomycin resistance protein
3017	6517	2925541	2926704	1164	gp:VCAJ10968_1	Vibrio cholerae JS1569 nptA	28.5	68.9	382	sodium dependent phosphate pump
3018	6518	2927546	2926707	840	sp:PHZC_PSEAR	Pseudomonas aureofaciens 30-84 phzC	38.8	56.4	289	phenazine biosynthesis protein
3019	6519	2928283	2927651	633						
3020	6520	2928318	2927551	768	gp:SCE8_16	Streptomyces coelicolor A3(2) SCE8.16c	24.3	8.09	255	ABC transporter
3021	6521	2929237	2928302	936	sp:BCRA_BACLI	Bacillus licheniformis ATCC 9945A bcrA	36.9	66.3	309	ABC transporter ATP-binding protein
3022	6522	2929756	2929256	501	pir.C70629	Mycobacterium tuberculosis H37Rv Rv0413	47.6	68.5	168	mutator mutT protein
3023	6523	2929951	2931336	1386	pir:B70629	Mycobacterium tuberculosis H37Rv Rv0412c	35.0	70.2	423	hypothetical membrane protein
3024	6524	2931340	2932371	1032	sp:GLNH_BACST	Bacillus stearothermophilus NUB36 glnH	31.5	64.8	270	glutamine-binding protein precursor
3025	6525	2932577	2934829	2253	pir:H70628	Mycobacterium tuberculosis H37Rv Rv0410c pknG	41.2	63.5	805	serine/threonine kinase
3026	6526	2933398	2932652	747						
3027	6527	2938403	2939767	1365	sp. ADRO_BOVIN	Bos taurus	37.2	8.79	457	ferredoxin/ferredoxin-NADP reductase
3028	6528	2939907	2940452	546	Sp:ELAA_ECOLI	Escherichia coli K12 elaA	34.0	60.3	156	acetyltransferase (GNAT) family
3029	6259	2941508	2940447	1062						
3030	6530	2942500	2941472	1029						
3031	6531	2943007	2942609	399						
3032	6532	2944205	2943012	1194	sp:PURT_BACSU	Bacillus subtilis 168 purT	59.1	82.6	379	phosphoribosylglycinamide formyltransferase
3033	6533	2946526	2945639	888						

										$\overline{}$				1	T			$-\tau$		7
5			related)	related)	n sensor	o.		thetase			ne protein	e aldolase			yltransferase					
10		Function	insertion element (IS3 related)	insertion element (IS3 related)	two-component system sensor histidine kinase	transcriptional regulator		adenylosuccinate synthetase	hypothetical protein		hypothetical membrane protein	fructose-bisphosphate aldolase	hypothetical protein	methyltransferase	orotate phosphoribosyltransferase	hypothetical protein	3-mercaptopyruvate sulfurtransferase			
15		Matched length (a.a.)	295	89	349	218		427	204		359	344	304	182	174	250	294			
20		Similarity (%)	90.9	84.3	51.3	65.6		95.3	59.3		100.0	100.0	100.0	91.2	65.5	0.09	56.1			
		Identity (%)	77.6	- 67.4	22.4	31.7		89.7	34.3		100.0	99.7	100.0	76.9	39.1	27.6	29.6			
25 3 	ontinuea)	s gene	lutamicum	lutamicum	noviolaceus	(36 degU		гA	erculosis		glutamicum 59 ORF3	glutamicum 59 fda	glutamicum 59 ORF 1	berculosis	i pyrE	berculosis	ST			
30	Table 1 (continued)	Homologous gene	Corynebacterium glutamicum orf2	Corynebacterium glutamicum orf1	Streptomyces thermoviolaceus opc-520 chiS	Bacillus brevis ALK36 degU		Corynebacterium ammoniagenes purA	Mycobacterium tuberculosis H37Rv Rv0358		Corynebacterium glutamicum AS019 ATCC 13059 ORF3	Corynebacterium glutamicum AS019 ATCC 13059 fda	Corynebacterium glutamicum AS019 ATCC 13059 ORF1	Mycobacterium tuberculosis H37Rv Rv0380c	Pyrococcus abyssi pyrE	Mycobacterium tuberculosis H37Rv Rv0383c	Homo sapiens mpsT		-	
35			25.5	S.F.	कें के			Co	₹£			28	3 %	ΣÏ	<u>e</u>	ΣÏ	<del> </del>			$\dashv$
40		db Match	pir.S60890	pir.S60889	gp:AB016841_1	sp.DEGU_BACBR		gp:AB003160_1	pir:G70575		sp:YFDA_CORGL	pir:S09283	gp:CGFDA_1	pir.G70833	gp:AF058713_1	pir. B70834	sp:THTM_HUMAN			
		ORF (bp)	894	267	1140	618	225	1290	759	264	1167	1032	951	618	552	972	852	720	279	399
<b>45</b>		Terminal (nt)	2946698	2947620	2948049	2949265	2950431	2950434	2952691	2952972	2952975	2954241	2955523	2956830	2957485		2959520	2960468	2962730	2963198
50		Initial (nt)	2947591	2947886	2949188	2949882	2950207	2951723	2951933	2952709		2955272	2956473	2957447	2958036		2960371	2961187	2963008	2963596
		SEQ NO.	6534	6535	6536	6537	6538	6239	6540	6541	6542	6543	6544	6545	6546		6548	6549		6551
55		SEQ NO.	3034	3035	3036	3037	3038	3039	3040	3041	3042	3043	3044	3045	3046	3047	3048	3049	3050	3051

5		Function	virulence factor	virulence factor	virulence factor	sodium/glutamate symport carrier protein	cadmium resistance protein	cation efflux system protein (zinc/cadmium)	monooxygenase or oxidoreductase or steroid monooxygenase	alkanal monooxygenase alpha chain		cystathionine gamma-lyase	bacterial regulatory protein, lacl family	rifampin ADP-ribosyl transferase	rifampin ADP-ribosyl transferase	hypothetical protein	hypothetical protein	oxidoreductase
15		fatched length (a.a.)	59 vir	200 vir	132 vir	489 pr	108 ca	283 ca	476 m	399 al		375 cy	184 ta	89	56 rri	361 h	204 h	386 0)
			- 5	50	<u>~</u>	4	Ĭ	2	4	Б		8	_		<u> </u>	6	2	۳ —
20		Similarity (%)	82.0	55.0	63.0	54.8	71.3	63.3	45.4	47.4		62.4	67.9	65.2	87.5	56.2	64.7	9.09
		Identity (%)	76.0	38.0	62.0	24.7	37.0	23.7	22.5	21.1		36.5	40.2	49.4	73.2	30.5	33.8	31.9
25	_						ပ			biont	·		(2)	(2)	(2)	S	S	S
30	Table 1 (continued)	Homologous gene	Pseudomonas aeruginosa ORF24222	Pseudomonas aeruginosa ORF23228	Pseudomonas aeruginosa ORF25110	Synechocystis sp. PCC6803 sin625	Staphylococcus aureus cadC	Pyrococcus abyssi Orsay PAB0462	Rhodococcus rhodochrous IFO3338	Kryptophanaron affredi symbiont luxA		Escherichia coli K12 metB	Streptomyces coelicolor A3(2) SC1A2.11	Streptomyces coelicator A3(2) SCE20.34c arr	Streptomyces coelicolor A3(2) SCE20.34c arr	Mycobacterium tuberculosis H37Rv Rv0837c	Mycobacterium tuberculosis H37Rv Rv0836c	Mycobacterium tuberculosis H37Rv Rv0385
33					20	0, 0,	+	<del> </del>		1								
40		db Match	GSP:Y29188	GSP:Y29182	GSP:Y29193	pir:S76683	Sp.CADF_STAAU	pir.H75109	gp:AB010439_1	sp:LUXA_KRYAS		Sp:METB_ECOLI	gp:SC1A2_11	gp:SCE20_34	gp:SCE20_34	pir.E70812	pir:D70812	pir.D70834
		ORF (bp)	177	762	396	1347	387	858	1170	1041	762	1146	567	240	183	1125	732	1179
45		Terminal (nt)	2964434	2965837	2965583	2966458	2968789	2969808	2971003	2972057	2971338	2972060	2973230	2974200	2974382	2975591	2976360	2977774
50		Initial (nt)	2964258	2965076	2965188	2967804	2968403		2969834	2971017	2972099	2973205		2973961	2974200	2974467	2975629	2976596
		SEQ NO.	6552	6553	6554	6555	6556	6557	6558	6559	6560	6561	6562	6563	6564	6565	6566	6567
55		SEQ NO.		3053	3054	3055	3056	3057	3058	3059	3060	3061	3062	3063	3064	3065	3066	3067

																	-		
5		c	o acid			regulatory	lase	on regulator	laJ	factor grpE ATPase domain serone DnaK	лаК	ne protein	ne ne nucleosidase			ation protein			sse
10		Function	N-carbamoyl-D-amino acid amidohydrolase		hypothetical protein	novel two-component regulatory system	aldehyde dehydrogenase	heat shock transcription regulator	heat shock protein dnaJ	nucleotide exchange factor grpE protein bound to the ATPase domain of the molecular chaperone DnaK	heat shock protein dnaK	hypothetical membrane protein	5'-methylthioadenosine nucleosidase and S- adenosylhomocysteine nucleosidase			chromosome segregation protein			alcohol dehydrogenase
15		Matched length (a.a.)	275		289	108	507	135	397	212	618	338	195			1311			334
<b>20</b>		Similarity (%)	67.3		55.4	44.0	90.3	70.4	80.1	66.5	93.8	79.0	0.09			48.4			81.7
		Identity (%)	32.0		. 28.0	38.0	9.69	47.4	56.7	38.7	99.8	42.6	27.2			18.9			20.0
25 :	nlinuea)	gene	m Delta H		olor A3(2)	nse carR	ropolis thcA	G hspR	erculosis aJ	color grpE	um MJ-233	color A3(2)	HP0089 mtn			ces pombe			mophilus
30	Table 1 (continued)	Homologous gene	Methanobacterium thermoautotrophicum Delta H MTH1811		Streptomyces coelicolor A3(2) SC4A7.03	Azospirillum brasilense carR	Rhodococcus erythropolis thcA	Streptomyces albus G hspR	Mycobacterium tuberculosis H37Rv RV0352 dnaJ	Streptomyces coelicolor grpE	Brevibacterium flavum MJ-233 dnaK	Streptomyces coelicolor A3(2) SCF6.09	Helicobacter pylori HP0089 mtn			Schizosaccharomyces pombe cut3		•	Bacillus stearothermophilus DSM 2334 adh
35			Z ÷ Z	<u> </u>	S Str	₹.	호	रु	<b> </b>	<del> </del>	के के	क छ	Ĭ						
40		db Match	pir.869109		gp:SC4A7_3	GP:ABCARRA_2	orf:2104333D	gp.SAU43299_2	sp:DNAJ_MYCTU	sp.GRPE_STRCO	gsp:R94587	gp:SCF6_8	sp.PFS_HELPY			sp:cUT3_SCHPO			sp:ADH2_BACST
•		ORF (bp)	798	243	1134	330	1518	438	1185	636	1854	1332	633	1200	885	3333	636	1485	1035
45		Terminal (nt)	2977847	2978979	2980115	2981216	2980181		1	2983887	2984544	2988164	2988214	2988846	2992602	2989954	2993286	2993921	2995747
50		Initial (nt)	2978644	2978737	2978982	2980887	298169R	2982460		2984522	2986397	2986833	2988846	2990045		2993286	2993921	2995405	
		SEQ NO.	6568	6269		6571	6577	6573	6574	6575	6576	6577	6578	6579	6580	6581	6582	6583	6584
55		SEQ NO.		3069		3071	2070			3075	3076	3077	3078	3079	3080	3081	3082	3083	3084

ammonia monooxygenase

161

76.4

39.1

Pseudomonas putida DSMZ ID 88-260 amoA

gp:PPAMOA\_1

522 321

3009710

3102 | 6602 | 3010231 |

486

hypothetical protein

8

66.3

50.0

Streptomyces coelicolor A3(2) SCE68.10

gp:SCE68\_10

366

3009607

3009242

6601

3101

					_												
5	r.					ne protein			ferase, subunit	ferase small	hosphosulfate	ductase	-NADP				take protein y
10	Function					hypothetical membrane protein	hypothetical protein		sulfate adenylyltransferase, subunit 1	sulfate adenylyltransferase small chain	phosphoadenosine phosphosulfate reductase	ferredoxin-nitrate reductase	ferredoxin/ferredoxin-NADP reductase	huntingtin interactor			alkylphosphonate uptake protein and C-P lyase activity
15	Matched length (a.a.)					301	252		414	308	212	205	487	144			142
20	Similarity (%)					70.1	53.2		78.3	70.1	64.2	65.5	61.4	59.7			59.9
	Identity (%)					43.5	32.5		47.3	46.1	39.2	34.5	30.8	32.6			26.8
25 D							(3(5)		_			7942	9				3
35 Table 1 (continued)	Homologous gene					Bacillus subtilis ytnM	Streptomyces coelicolor A3(2) SC7A8.10c		Escherichia coli K12 cysN	Escherichia coli K12 cysD	Bacillus subtilis cysH	Synechococcus sp. PCC 7942	Saccharomyces cerevisiae FL200 arh1	Homo sapiens hypE			Escherichia coli K12 phnB
40	db Match					pir.F69997	gp:SC7A8_10		sp:CYSN_ECOLI	sp.CYSD_ECOLI	sp:CYH1_BACSU	Sp. NIR_SYNP7	sp:ADRO_YEAST	prf:2420294J			sp:PHNB_ECOL!
						1	<del></del>	1.0		$\overline{}$		_			_	**	
	ORF (bp)	216	207	189	261	927	723	915	1299	912	693	1683	1371	1083	237	534	414
45	Terminal (nt)	2997366	2997481	2997876	2997963	2998528	2999478	3002426	3000241	3001542	3002453	3003480	3006915	3008376	3008453	3009303	3008749
50	Initial (nt)	2997151	2997687	2997688	2998223	2999454	3000200	3001512		3002453	3003145	3005162	3005545	3007294	3008689	3008770	3009162
	SEQ NO.	6585	9859	6587	6588	6839	6590	6591	6592	6593	6594	6595	6596	6597	6598	6299	0099
55	SEQ NO.	3085	3086	3087	3088	3089	3090	3091	3092	3093	3094	3095	3096	3097	3098	3099	3100

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

Te		Termina	1 _	ORF	db Match	Homologous gene	Identity	Similarity	Matched	Function
(nt) (nt) (bp)	(nt) (bp)	(dq)					(%)	(%)	(a.a.)	
6605 3010989 3011273 285 SP:YTZ3_AGRVI	3010989 3011273 285 S	285 S	တ	SP:YTZ3_AGR\	=	Agrobacterium vitis ORFZ3	41.0	58.0	89	hypothetical protein
6606 3011805 3011242 564	3011805 3011242		564							
6607 3012809 3011808 1002 sp.YGB7_ALCEU	3012809 3011808 1002	1002				Alcaligenes eutrophus H16 ORF7	26.1	57.9	337	hypothetical protein
6608 3013798 3013106 693 gp:HIU68399_3	3013798 3013106 693	693		gp:HIU68399_3		Haemophilus influenzae hmcB	35.7	64.8	199	ABC transporter
6609 3014550 3013837 714 gp:HIU68399_3	3014550 3013837 714	714		gp:HIU68399_3		Haemophilus influenzae hmcB	39.3	73.0	211	ABC transporter
6610 3014616 3015824 1209 pir.A69778	3014616 3015824 1209	1209	_	pir:A69778	4	Bacillus subtilis ydeG	30.8	8.79	416	metabolite transport protein homolog
6611 3015469 3014648 822	3014648		822		- 1					
6612 3016238 3016924 687	3016238 3016924		289							
6613 3017149 3015827 1323 sp.DAPE_ECOLI	3017149 3015827 1323	1323				Escherichia coli K12 msgB	21.5	48.5	466	succinyl-diaminopimelate desuccinylase
6614 3017316 3019220 1905	3017316 3019220	Н	1905		1 1					
6615 3017539 3018312 774	3017539 3018312		774							
6616 3018181 3017420 762	3018181 3017420		762							
6617 3019076 3018123 954 GPU:DCA297422_	3019076 3018123 954 G	954 G	9+	GPU:DCA297422_ 1		Daucus carota	33.0	46.0	114	dehydrin-like protein
6618 3020609 3019542 1068 sp:MALK_ECOLI	3020609 3019542 1068	1068				Escherichia coli K12 malK	24.9	50.1	373	maltose/maltodextrin transport ATP- binding protein
6619 3021202 3020561 642	3021202 3020561		642							
6620 3021825 3021208 618 gp:AF036485_6	3021825 3021208 618	618		gp:AF036485_6		Lactococcus lactis Plasmid pNZ4000 Orf-200 cbiM	30.2	9.79	179	cobalt transport protein
6621 3022928 3022113 816 sp.FRP_VIBHA	3022113 816	816		sp:FRP_VIBHA		Vibrio harveyi MAV frp	37.2	71.4	231	NADPH-flavin oxidoreductase
6622 3023900 3022998 903 sp:IUNH_CRIFA	3023900 3022998 903	903		sp:IUNH_CRIFA		Crithidia fasciculata iunH	28.4	59.3	317	inosine-uridine preferring nucleoside hydrolase
6623 3024379 3025353 975 gp:SCE20_8	3024379 3025353 975	975		gp:SCE20_8		Streptomyces coelicolor A3(2) SCE20.08c	31.2	59.4	276	hypothetical membrane protein
6624 3025552 3026139 588 sp.3MG1_ECOLI	3025552 3026139 588	588		sp:3MG1_ECOLI		Escherichia coli K12 tag	50.3	78.8	179	DNA-3-methyladenine glycosylase
6625 3027299 3026142 1158 sp.HMPA_ALCEU	3027299 3026142 1158	1158				Alcaligenes eutrophus H16 fhp	33.5	63.8	406	flavohemoprotein

															_						
5		-				nator or beta- ulatory protein		sidase		sidase	erase			e protein		genase	hate		-		ninidase
10		Function		oxidoreductase		transcription antiterminator or beta- glucoside positive regulatory protein		6-phospho-beta-glucosidase		6-phospho-beta-glucosidase	aspartate aminotransferase		transposase (ISCg2)	hypothetical membrane protein		UDP-glucose dehydrogenase	deoxycytidine triphosphate deaminase		hypothetical protein		beta-N-Acetylglucosaminidase
15		Matched length (a.a.)		210		192		167		99	402		401	399		442	188		229		410
20		Similarity (%)		63.8		69.3		59.9		78.8	80.9		100.0	70.2		72.2	72.3		59.4		58.1
		Identity (%)		34.8		28.1		43.7		43.9	53.7		100.0	33.6		40.5	43.6		30.6		28.5
25	inued)	ene		or A3(2)		glC		ım B6405		ım B6405	atus aat		amicum	or A3(2)		rkpK	lcd		or A3(2)		violaceus
30	Table 1 (continued)	Homologous gene	-	Streptomyces coelicolor A3(2) mmyQ		Escherichia coli K12 bglC	,	Clostridium longisporum B6405 abgA		Clostridium longisporum B6405 abgA	Methylobacillus flagellatus aat		Corynebacterium glutamicum ATCC 13032 tnp	Streptomyces coelicolor A3(2) SCQ11.10c		Sinorhizobium meliloti rkpK	Escherichia coli K12 dcd		Streptomyces coelicolor A3(2) SCC75A.16c		Streptomyces thermoviolaceus nagA
35											Σ		OA	S S		S	m_		တ်တ		σĒ
40		db Match		gp:SCO276673_18		sp:BGLG_ECOLI		sp:ABGA_CLOLO		sp:ABGA_CLOLO	gp:L78665_2		gp:AF189147_1	gp:SCQ11_10		prf.2422381B	sp.DCD_ECOLI		gp:SCC75A_16		gp:AB008771_1
		ORF (bp)	603	624	156	591	279	360	381	240	1257	30	1203	1257	183	1317	567	237	771	1689	1185
45		Terminal (nt)	3028163	3028891	3029033	3028884	3029782	3029702	3030535	3030101	3031979	3032348	3033863	3035437	3034105	3035440	3036845	3037911	3038942	3038993	3040748
50		Initial (nt)	3027561	3028268	3028878	3029474	3029504	3030061	3030155	3030340	3030723	3032647	3032661	3034181	3034287	3036756	3037411	3037675	3038172	3040681	3041932
		SEQ NO.	6626	6627	6628	6629	6630	6631	6632	6633	6634	6635	6636	6637	6638	6639	6640	6641	6642	6643	6644
		go€	26	27	28	29	8	31	32	8	34	35	88	37	38	39	40	41	42	143	144

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

	Function			hypothetical protein			hypothetical membrane protein	acyltransferase or macrolide 3-O-acyltransferase		hypothetical membrane protein		hexosyltransferase	methyl transferase	phosphoenolpyruvate carboxykinase (GTP)	C4-dicarboxylate transporter	hypothetical protein	hypothetical protein	mebrane transport protein	
	Matched length (a.a.)			1416			363	408		529		369	251	601	332	241	207	768	
	Similarity (%)			49.4			47.1	51.0		54.8		79.1	73.3	78.5	52.7	67.2	85.0	72.3	
	Identity (%)			29.6			24.8	27.7		31.2		53.4	58.6	54.7	24.4	35.7	69.1	42.3	
Table 1 (continued)	Homologous gene			Mycobacterium leprae MLCB1883.13c			Mycobacterium leprae MLCB1883.05c	Streptomyces sp. acyA		Mycobacterium leprae MLCB1883.04c		Mycobacterium tuberculosis H37Rv Rv0225	Mycobacterium tuberculosis H37Rv Rv0224c	Neocallimastix frontalis pepck	Pyrococcus abyssi Orsay PAB2393	Escherichia coli K12 yggH	Mycobacterium tuberculosis H37Rv Rv0207c	Mycobacterium tuberculosis H37Rv Rv0206c mmpL3	
	db Match			gp:MLCB1883_7			gp:MLCB1883_4	pir.JC4001		gp:MLCB1883_3		pir.G70961	pir.F70961	sp.PPCK_NEOFR	pir.E75125	sp:YGGH_ECOLI	pir:E70959	pir.C70839	
	ORF (bp)	444	201	3129	621	195	903	1068	708	1422	699	1137	177	1830	1011	765	705	2316	1422
	Terminal (nt)	3042437	3042703	3045788	3043022	3045990	3048048	3046122	3047197	3049479	3051190	3049456	3051964	3052062	3055769	3056631	3057317	3059643	3058096
	Initial (nt)	3041994	3042503	3042660	3043642	3045796	3047146	3047189	3047904	3048058	3050522	3050592	3051194	3053891	3054759	3055867	3056613	3057328	3059517
	SEQ NO.	6645	6646	6647	6648	6649	6650	6651	6652	6653	6654	6655	9999	6657	6658	6659	0999	6661	9999
	SEQ NO.	3145	3146	3147	3148	3149	3150	3151	3152	3153	3154	3155	3156	3157	3158	3159	3160	3161	3162

5	Function	hypothetical membrane protein	hypothetical membrane protein	propionyl-CoA carboxylase complex B subunit	polyketide synthase	acyl-CoA synthase	hypothetical protein		major secreted protein PS1 protein precursor			antigen 85-C	hypothetical membrane protein	nodulation protein	hypothetical protein	hypothetical protein		phosphatidic acid phosphatase
15	Matched tength (a.a.)	364 h	108	523	1747	592	319		657			331	. 299	295	168	656		170
20	Similarity (%)	62.9	69.4	76.9	54.2	62.3	67.4		99.5			62.5	61.2	51.5	75.0	74.7		56.5
	Identity (%)	29.1	34.3	49.7	30.2	33.5	39.8		98.6			36.3	37.5	27.1	51.2	55.6		28.2
25 (penuiju	gene	rculosis	rculosis	olor A3(2)	aeus eryA	s BCG	rculosis		utamicum um) ATCC			erculosis 3C fbpC	erculosis	nodans	erculosis	erculosis		is ATCC
30 Table 1 (continued)	Homologous gene	Mycobacterium tuberculosis H37Rv Rv0204c	Mycobacterium tuberculosis H37Rv Rv0401	Streptomyces coelicolor A3(2) pccB	Streptomyces erythraeus eryA	Mycobacterium bovis BCG	Mycobacterium tuberculosis H37Rv Rv3802c		Corynebacterium glutamicum (Brevibacterium flavum) ATCC 17965 cop1			Mycobacterium tuberculosis ERDMANN RV0129C fbpC	Mycobacterium tuberculosis H37Rv Rv3805c	Azorhizobium caulinodans ORS571 noeC	Mycobacterium tuberculosis H37Rv Rv3807c	Mycobacterium tuberculosis H37Rv Rv3808c		Bacillus licheniformis ATCC 9945A bcrC
35		ΣI	≥I		<u> </u>		21						21					
40	db Match	pir:A70839	pir:H70633	gp:AF113605_1	sp:ERY1_SACER	prf:2310345A	pir.F70887		sp.CSP1_CORGL			sp:A85C_MYCTU	pir.A70888	sp:NOEC_AZOCA	pir.C70888	pir.D70888		sp:BCRC_BACLI
	ORF (bp)	1083	363	1548	4830	1788	927	498	1971	1401	219	1023	2058	966	504	1968	1494	477
45	Terminal (nt)	3060733	3061095	3061380	3062951	3068143	3070214	3071147	3071650	3075447	3073857	3075540	3076715	3078853	3079848	3080344	3083960	3083935
50	Initial (nt)	3059651	3060733	3062927	3067780	3069930	3071140	3071644	3073620	3074047	3074075	3076562	3078772	3079848	3080351	3082311	3082467	3084411
	SEQ NO.	6663	6664	6665	9999	6667	8999	6999	6670	6671	6672	6673	6674	6675	9299	6677	8678	6679
55	SEQ		3164	3165	3166	3167	3168	3169	3170	3171	3172	3173	3174	3175	3176	3177	3178	3179

	-
5	
10	-
15	
20	
25	(panu
30	Table 1 (continued)
35	
40	
45	
50	

Function			dimethylaniline monooxygenase (Noxide-forming)		UDP-galactopyranose mutase	hypothetical protein	glycerol kinase	hypothetical protein	acyltransferase	seryl-tRNA synthetase	transcriptional regulator, GntR family or fatty acyl-responsive regulator	hypothetical protein	hypothetical protein		2,3-PDG dependent phosphoglycerate mutase		nicotinamidase or pyrazinamidase	
Matched length (a.a.)			377		377	629	499	279	261	419	235	356	113		218		460	
Similarity (%)			50.4		72.9	47.8	78.8	70.3	72.0	87.6	61.7	61.2	7.67		62.8		50.9	
Identity (%)			24.4		43.2	29.6	51.7	41.6	46.7	70.2	27.72	32.6	46.0		37.2		27.4	
Homologous gene			Sus scrofa fmo1		Escherichia coli K12 glf	Mycobacterium tuberculosis H37Rv Rv3811 csp	Pseudomonas aeruginosa ATCC 15692 glpK	Mycobacterium tuberculosis H37Rv Rv3813c	Mycobacterium tuberculosis H37Rv Rv3816c	Mycobacterium tuberculosis H37Rv	Escherichia coli K12 farR	Mycobacterium tuberculosis H37Rv Rv3835	Mycobacterium tuberculosis H37Rv Rv3836		Amycolatopsis methanolica pgm		Mycobacterium smegmatis pzaA	
db Match			sp:FMO1_PIG		sp.GLF_ECOLI	pir:G70520	sp:GLPK_PSEAE	pir.A70521	pir.D70521	gsp:W26465	sp:FARR_ECOLI	pir.H70652	pir.A70653		gp:AMU73808_1		prf:2501285A	
ORF (bp)	777	510	1302	612	1203	2049	1527	834	876	1266	714	1113	342	66	699	630	1143	729
Terminal (nt)	3084424	3085218	3087048	3088276	3087101	3090664	3090760	3092342	3093175	3094078	3096287	3097423	3097764	3097780	3097904	3099454	3100698	3101426
Initial (nt)	3085200	3085727	3085747	3087665	3088303	3088616	3092286	3093175	3094050	3095343	3095574	3096311	3097423	3097878	<del></del>	3098825	3099556	3100698
SEQ NO.	9899	6681	6682	6683	6684	6685	9899	6687	6688		0699	6691	6692	6693	6694	6695	+	6697
SEQ NO.	3180	3181	3182	3183	3184	3185	3186	3187	3188	3189	3190	3191	3192	3193	3194	3195	3196	3197

	_																				
5		Function	transcriptional regulator				hypothetical protein	glucan 1,4-alpha-glucosidase		glycerophosphoryl diester phosphodiesterase	gluconate permease			kinase	L-lactate dehydrogenase	hypothetical protein	hydrolase or haloacid dehalogenase-like hydrolase	tein	transcription activator or transcriptional regulator GntR family	esterase	shikimate transport protein
15		pe q												pyruvate kinase				efflux protein		phosphoesterase	
		Matched length (a.a.)	380				107	432		259	456			491	314	526	224	188	221	255	422
20		Similarity (%)	57.1				81.3	55.3		54.1	71.9			47.7	99.7	64.8	58.5	67.6	57.0	68.6	74.4
		Identity (%)	31.6				43.9	28.7		29.0	37.3			25.5	99.7	33.5	32.1	39.9	27.6	47.8	37.9
25	inued)	ane	or A3(2)				lae	isiae						amicum	ı lctA	culosis	lor A3(2)	ORF1	AG1655	culosis	shiA
30	Table 1 (continued)	Homologous gene	Streptomyces coelicolor A3(2) SC6G4.33				Streptomyces lavendulae ORF372	Saccharomyces cerevisiae S288C YIR019C sta1		Bacillus subtilis glpQ	Bacillus subtilis gntP			Corynebacterium glutamicum AS019 pyk	Brevibacterium flavum lctA	Mycobacterium tuberculosis H37Rv Rv1069c	Streptomyces coelicolor A3(2) SC1C2.30	Brevibacterium linens ORF1 tmpA	Escherichia coli K12 MG1655 glcC	Mycobacterium tuberculosis H37Rv Rv2795c	Escherichia coli K12 shiA
<i>35</i> <i>40</i>		db Match	gp:SC6G4_33				pir.B26872	sp:AMYH_YEAST		sp.GLPQ_BACSU [6	Sp.GNTP_BACSU			sp:KPYK_CORĠL	gsp:Y25997	pir.C70893	gp:SC1C2_30	gp:AF030288_1	sp:GLCC_ECOLI	pir:B70885	sp:SHIA_ECOLI
		ORF (bp)	1035	120	552	870	327	1314	918	819	1389	642	159	1617	942	1776	636	543	693	786	1299
45		Terminal (nt)	3102768	3101744	3102079	3103763	3104252	3105719	3106053	3106951	3109519	3108823	3110003	3110464	3112449	3115394	3116042	3116621	3117332	3118121	3119582
50		Initial (nt)	3101734	3101863	3102630	3102894	3103926	3104406	3106970	3107769	3108131		3109845	3112080	3113390	3113619	3115407	3116079	3116640	3117336	3118284
		SEQ NO.	6698	6699	6700	6701	6702	6703	6704	6705	6706	+	6708	-	6710	6711	6712	6713	6714	6715	6716
55		SEQ NO.	3198	3199	3200	3201	3202	3203	3204	3205	3206	3207	3208	3209	3210	3211	3212	3213	3214	3215	3216

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

	Function	L-lactate dehydrogenase or FMN-dependent dehydrogenase	-	immunity repressor protein			phosphatase or reverse transcriptase (RNA-dependent)		peptidase or IAA-amino acid hydrolase		peptide methionine sulfoxide reductase	superoxide dismutase (Fe/Mn)	transcriptional regulator	multidrug resistance transporter				hypothetical protein	membrane transport protein	transcriptional regulator	two-component system response regulator	
	Matched length (a.a.)	376		55			569		122		210	164	292	384				216	447	137	212	
	Similarity (%)	68.9		80.0			51.3		63.1		69.1	92.7	65.8	49.0				64.8	59.3	65.0	75.5	
	Identity (%)	40.4		45.5			29.5		36.9		47.6	82.3	32.5	23.4			Ì	33.8	27.3	37.2	50.9	
Table 1 (continued)	Homologous gene	Neisseria meningitidis IIdA		Bacillus phage phi-105 ORF1			Caenorhabditis elegans Y51B11A.1		Arabidopsis thaliana ill1		Escherichia coli B msrA	Corynebacterium pseudodiphtheriticum sod	Bacillus subtilis gltC	Corynebacterium glutamicum tetA				Mycobacterium tuberculosis H37Rv Rv3850	Streptomyces cyanogenus lanJ	Bacillus subtilis 168 yxaD	Corynebacterium diphtheriae chrA	
	db Match	prf:2219306A		sp:RPC_BPPH1			gp:CELY51B11A_1		sp.ILL1_ARATH		sp:PMSR_ECOLI	pir:140858	sp:GLTC_BACSU	gp:AF121000_10				pir.G70654	prf:2508244AB	sp:YXAD_BACSU	prf.2518330B	
	ORF (bp)	1215	405	312	138	711	1617	546	402	150	651	909	924	1134	1611	111	1521	633	1491	456	636	
	Terminal (nt)	3120879	3121313	3121909	3121992	3123932	3122556	3124341	3124897	3125492	3125495	3126991	3127494	3129739	3131395	3133030	3131508	3133747	3133778	3135752	3135856	
	Initial (nt)	3119665	3120909	3121598	6720 3122129	3123222	3124172	3124886	3125298	3125343	3126145	3126392	3128417	3128606	3129785	3132920	3133028	3133115	3135268	3135297	3136491	
	SEQ NO.	6717	6718	6719		6721	6722	6723	6724	6725	6726	6727	6728	6729	6730	6731	6732	6733	6734	6735	6736	
	SEQ NO. (DNA)	3217	3218	3219	3220	3221	3222	3223	3224	3225	3226	3227	3228	3229	3230	3231	3232	3233	3234	3235	3236	

5	Function			two-component system sensor histidine kinase	hypothetical protein	hypothetical protein	stage III sporulation protein	transcriptional repressor	transglycosylase-associated protein	hypothetical protein	hypothetical protein	RNA pseudouridylate synthase	hypothetical protein	hypothetical protein		bacterial regulatory protein, gntR family or glc operon transcriptional activator	hypothetical protein	hypothetical protein
15	<b>v</b> _ v	+	1	histic	hypo	hypo	stag	trans	tran				hyp	hyp				
	Matched length (a.a.)			408	48	277	265	192	87	296	314	334	26	42		109	488	267
20	Similarity (%)			64.5	79.2	59.2	53.6	6.09	71.3	9.69	73.9	51.2	0.99	75.0		56.0	48.2	78.7
	Identity (%)			30.2	45.8	30.0	26.0	32.3	34.5	41.2	38.5	28.4	61.0	71.0		30.3	26.0	48.3
<i>25</i>				ae	3(2)	3(2)		is	655	sis	655	305		Đ		655		Sis
Se Table 1 (continued)	Homologous gene			Corynebacterium diphtheriae chrS	Streptomyces coelicolor A3(2) SCH69.22c	Streptomyces coelicolor A3(2) SCH69.20c	Bacillus subtilis spolliJ	Mycobacterium tuberculosis H37Rv Rv3173c	Escherichia coli K12 MG1655 tag1	Mycobacterium tuberculosis H37Rv Rv2005c	Escherichia coli K12 MG1655. yhbW	Chlorobium vibrioforme ybc5	Chlamydia pneumoniae	Chlamydia muridarum Nigg TC0129		Escherichia coli K12 MG1655 glcC	Streptomyces coelicolor SC4G6.31c	Mycobacterium tuberculosis H37Rv Rv2744c
35			_	ਠੌਂ ਓ	20.00	8 8		ΣΪ						ō ¥	-		מ מ	
40	db Match			prf:2518330A	gp:SCH69_22	gp:SCH69_20	sp:SP3J_BACSU	pir.C70948	sp:TAG1_ECOLI	sp:YW12_MYCTU	Sp:YHBW_ECOLI	Sp.YBC5_CHLVI	<del></del>	PIR:F81737		sp:GLCC_ECOLI	gp:SC4G6_31	sp:35KD_MYCTU
	ORF (bp)	639	588	1311	150	822	1302	639	261	903	987	996	273	141	207	363	1416	873
45	Terminal (nt)	3137558	3138471	3136593	3138481	3138634	3140952	3140885	3141709	3142454	3143496	3145626	3146841	3147230	3151369	3151842	3153828	3153894
50	Initial (nt)	3136920	3137884	3137903	3138630	3139455	3139651	3141523	3141969	3143356	3144482	3144661		3147090	3151575	£	3152413	3154766
	SEQ NO.	6737	6738	6239	6740	6741	6742	6743	6744	6745	6746	6747	6748		6750	6751	6752	6753
55	SEQ NO. (DNA)	3237	3238	3239	3240	3241	3242	3243	3244	3245	3246	3247	3248	3249	3250	3251	3252	3253

		Γ	Τ	Τ	Τ		Ţ	Г	Τ	T	Τ	$\top$	Τ	Т	T -	Ţ	<u> </u>	Τ	<u> </u>	T	Ι	$\overline{\Box}$
5				-				tein				olvase					agment		sphate dogene)		sporting Insporting	
10	Function						methyltransferase	nodulin 21-related protein				transposon tn501 resolvase		ferredoxin precursor	hypothetical protein	transposase	transposase protein fragment TnpNC		glyceraldehyde-3-phosphate dehydrogenase (pseudogena)	lipoprotein	copper/potassium-transporting ATPase B or cation transporting ATPase (E1-E2 family)	
15	Matched length (a.a.)						217	241				56		62	55	27	46		38	180	717	
20	Similarity (%)						58.1	55.2				92.9		98.4	85.5	84.0	90.0		84.2	59.4	73.4	
	Identity (%)						32.3	26.1				48.2		90.3	47.3	81.0	84.0		63.2	32.2	45.8	
Table 1 (continued)	us gene						licolor A3(2)					uginosa TNP5		a erythraea fer	licolor A3(2)	glutamicum	glutamicum		я дар	PCC6803	gidus AF0152	
7able 1 (	Homologous gene						Streptomyces coelicolor A3(2) SCD35.11c	soybean NO21				Pseudomonas aeruginosa TNP5		Saccharopolyspora erythraea fer	Streptomyces coelicolor A3(2)	Corynebacterium glutamicum Tnp1673	Corynebacterium glutamicum		Pyrococcus woesei gap	Synechocystis sp. PCC6803 sII0788	Archaeoglobus fulgidus AF0152	
40	db Match						p:SCD35_11	sp:NO21_SOYBN				sp:TNP5_PSEAE		sp:FER_SACER	gp:SCD31_14	GPU:AF164956_8	GPU:AF164956_23		p:G3P_PYRWO	ir.S77018	ir.H69268	-
	ORF (bp)	153	1452	1068	249	309	711 g	720 s	204	378	186	216 s	483	321 s	333 g	111	162	1038	126 s	999	2217 p	171
45	Terminal (nt)	3154969	3155246	3156306	3157223	3157479	3158834	3159081	3160419	3161065	3161001	3160723	3161701	3161087	3161682	3162804	3162871	3163889	3162858	3163074	3163789	3166267
50	Initial (nt)	3154817	3156697	3157373	3157471	3157787	3158124	3159800	3160216	3160688	3160816	3160938	3161219	3161407	3162014	3162694	3162710	3162852	3162983	3163733	3166005	3166437
	SEQ NO. (a.a.)	6754	6755	6756	6757	6758	6229	6760	6761	6762	6763	6764	6765	99/9	2929	6768	6929	6770	6771	6772	6773	6774
55	SEQ NO. (DNA)	3254	3255	3256	3257	3258	3259	3260	3261	3262	3263	3264	3265	3266	3267	3268	3269	3270	3271	3272	3273	3274

5	Function		two-component system sensor histidine kinase		two-component response regulator or alkaline phosphatase synthesis transcriptional regulatory protein		laccase or copper resistance protein precursor A	thiol disulfide interchange protein (cytochrome c biogenesis protein)	quinone oxidoreductase (NADPH: quinone reductase)(seta- crystallin)		zinc-transporting ATPase (Zn(II)- translocating p-type ATPase			zinc-transporting ATPase (Zn(II)- translocating p-type ATPase	hypothetical protein		transposase	transposase
15	Matched length (a.a.)		301 4		233 o		630 PP	101	322 (		78 Z			606	72 h		73 t	70 t
20	Similarity (%)		71.4		72.1		47.9	63.4	60.9		66.7			68.5	54.0		73.0	77.0
	Identity (%)		37.5		43.4		26.7	31.7	31.4		37.2			39.8	45.0		58.0	75.0
30 Table 1 (continued)	us gene		12 baeS		ОР		ingae pv.	ponicum tlpA			PCC6803			.12 MG1655	K1 APE2572		glutamicum	glutamicum
Table 1	Homologous gene		Escherichia coli K12 baeS		Bacillus subtilis phoP		Pseudomonas syringae pv. tomato copA	Bradyrhizobium japonicum tlpA	Mus musculus qor		Synechocystis sp. atzN			Escherichia coli K12 MG1655 atzN	Aeropyrum pernix K1 APE2572		Corynebacterium glutamicum Tnp1673	Corynebacterium glutamicum Tnp1673
35	atch						-											
40	db Match		sp:BAES_ECOLI		sp:PHOP_BACSU		sp.COPA_PSESM	sp:TLPA_BRAJA	sp:QOR_MOUSE		sp:ATZN_SYNY3			sp:ATZN_ECOLI	PIR:E72491		GPU:AF164956_8	GPU:AF164956_8
	ORF (bp)	192	1197	828	756	672	1479	363	918	471	234	315	207	1875	390	309	216	258
45	Terminal (nt)	3167169	3166450	3168566	3167646	3169340	3170892	3171616	3171619	3173465	3173857	3174380	3174784	3176901	3175254	3177482	3177089	3177308
50	Initial (nt)	3166978	3167646	3167739	3168401	3168669	3169414	3171254	3172536	3172995	3173624	3174066	3174990	3175027	3175643	3177174	3177304	3177565
	SEQ NO.	6775	6776	5777	6778	6779	6780	6781	6782	6783	6784	6785	6786	6787	6788	6789	0629	6791
55	SEQ NO. (DNA)	3275	3276	3277	3278	3279	3280	3281	3282	3283	3284	3285	3286	3287	3288	3289	3290	3291

EP 1 108 790 A2

5	Function	(28)			ansport protein or e transporter			licase		itein L9	binding protein	itein S6		Ē	-	orotein	Ŀ	y protein, marR	u		ni	_	ABC transporter ATP-binding protein
10	Fun	transposase (IS1628)	thioredoxin		transmembrane transport protein or 4-hydroxybenzoate transporter		hypothetical protein	replicative DNA helicase		50S ribosomal protein L9	single-strand DNA binding protein	30S ribosomal protein S6		hypothetical protein		penicillin-binding protein	hypothetical protein	bacterial regulatory protein, marR family	hypothetical protein		hypothetical protein	hypothetical protein	ABC transporter A
15	Matched length (a.a.)	53	100		421		208	461		154	229	92		480		647	107	137	296		7.1	298	433
20	Similarity (%)	96.2	74.0		60.1		62.5	73.1		71.4	51.5	78.3		68.3		60.1	72.0	0:59	61.8		70.4	63.8	64.0
	Identity (%)	92.5	39.0		27.1		35.1	37.7		42.2	30.6	28.3		41.5		29.1	41.1	35.1	29.7		32.4	30.2	31.2
25 (panining) 1 and 1	ens gene	glutamicum pAG1 tnpB	(12 thi2		ıtida pcaK		(12 yaji	(12 dnaB		(12 RL9	(12 ssb	(12 RS6		megmatis		onA	berculosis	berculosis	sperculosis yofF		hgC	(12 yceA	(12 ybjZ
30 d	Homologous gene	Corynebacterium glutamicum 22243 R-plasmid pAG1 tnpB	Escherichia coli K12 thi2		Pseudomonas putida pcaK		Escherichia coli K12 yqil	Escherichia coli K12 dnaB		Escherichia coli K12 RL9	Escherichia coli K12 ssb	Escherichia coli K12 RS6		Mycobacterium smegmatis mc(2)155		Bacillus subtilis ponA	Mycobacterium tuberculosis H37Rv Rv0049	Mycobacterium tuberculosis H37Rv Rv0042c	Mycobacterium tuberculosis H37Rv Rv2319c yofF		Bacillus subtilis yhgC	Escherichia coli K12 yceA	Escherichia coli K12 ybjZ
<i>35</i>	db Match	gp:AF121000_8	sp:THI2_ECOLI		sp:PCAK_PSEPU		sp:YQJI_ECOLI	sp:DNAB_ECOLI		sp:RL9_ECOLI	sp:SSB_ECOL!	sp:RS6_ECOLI		gp:AF187306_1		sp:PBPA_BACSU	sp:Y0HC_MYCTU	pir:B70912	sp:Y0FF_MYCTU		sp:YHGC_BACSU	sp:YCEA_ECOLI	sp:YBJZ_ECOLI
	ORF (bp)	159 (	447	264	1344 8	159	576	1530	516	450	675	285	189	1458 g	882	2160	357 8	471	942 s	495	321	936	1263
<b>45</b> ,	Terminal (nt)	3177525	3178112	3178872	3180392	3180946	3180551	3181337	3183984	3183478	3183987	3184701	3185348	3185536	3188793	3187042	3189296	3190347	3191319	3191848	3191922	3192266	3193252
50	Initial (nt)	3177683	3178558	3178609	3179049	3181104	3181126	3182866	3183469	3183927	3184661	3184985	3185536	3186993	3187912	3189201	3189652	3189877	3190378	3191354	3192242	3193201	3194514
	SEQ NO.	6792	6793	6794	6795	9629	6797	6798	6429	9800	6801	6802	6803	6804	6805	6806	6807	6808	6899	6810	6811	6812	6813
55	SEQ NO.	3292	3293	3294	3295	3296	3297	3298	3299	3300	3301	3302	3303	3304	3305	3306	3307	3308	3309	3310	3311	3312	3313

5	uo	-binding protein					ng starvation	e-DNA				otein-cysteine	ogenase or tase ductase) or		t protein	ase [NAD] (malic	uconate kinase	ce protein	ce protein
10	Function	ABC transporter ATP-binding protein	hypothetical protein	hypothetical protein			DNA protection during starvation protein	formamidopyrimidine-DNA glycosylase	hypothetical protein			methylated-DNA-protein-cysteine S-methyltransferase	zinc-binding dehydrogenase or quinone oxidoreductase (NADPH:quinone reductase) or atginate tyase		membrane transport protein	malate oxidoreductase [NAD] (malic enzyme)	gluconokinase or gluconate kinase	teicoplanin resistance protein	teicoplanin resistance protein
15	Matched length (a.a.)	221	237	360			154	268	404			166	231		398	392	486	169	159
20	Similarity (%)	80.1	42.0	90.0			64.9	55.6	9.99			63.3	63.6		66.3	99.5	53.7	60.4	159.0
	identity (%)	48.9	18.0	77.8			37.7	28.4	47.5			38.0	33.3		26.4	99.7	24.5	27.8	27.0
25 (pent	90	31655	9090	losis			S	utM or	සු				ea pig) qor		ulosis	ssecola amicum)		vanZ	vanZ
Se Table 1 (continued)	Homologous gene	Escherichia coli K12 MG1655 ybjZ	Campylobacter jejuni Cj0606	Mycobacterium tuberculosis H37Rv Rv0046c			Escherichia coli K12 dps	Escherichia coli K12 mutM or fpg	Escherichia coli K12 rtcB			Homo sapiens mgmT	Cavia porcellus (Guinea pig) qor		Mycobacterium tuberculosis H37Rv Rv0191 ydeA	Corynebacterium melassecola (Corynebacterium glutamicum) ATCC 17965 malE	Bacillus subtilis gntK	Enterococcus faecium vanZ	Enterococcus faecium vanZ
35			Ö	ΣI					1				<u> </u>	-					1
40	db Match	sp:YBJZ_ECOLI	pir.E81408	pir:F70912			sp.DPS_ECOU	sp:FPG_ECOLI	sp.RTCB_ECOLI			SP:MGMT_HUMAN	sp:QOR_CAVPO		sp:YDEA_ECOLI	gp:AF234535_1	SP.GNTK BACSU	+	-
	ORF (bp)	069	1977	1089	909	1485	495	813	1149	1089	573	474	1011	E	1176	1176	1482	<del></del>	1-1
45	Terminal (nt)	3194514	3195210	3198500	3198582	3199202	3201260	3202712	3204100	3202979	3204728		3205222	3206756		3209454	3209705		
50	Initial (nt)	3195203	3197186	3197412	3199187	3200686	3201754	3201900	3202952		3204156	3205204	3206232	3206646		3208279	3211186		
	SEQ NO.	6814	6815	6816	6817	6818	6819	6820	6821	6822	_		6825	6826	<del></del>	6828	6829		
55	SEQ.		3315	_	3317	3318		3320	3321	3322	3373	3324	3325	3326	3327	3328	3320	3330	3331

adation sport protein or ransporter
family or pectin degradation repressor protein transmembrane transport protein or 4-hydroxybenzoate transporter
229
60.8
25.3
ysanthemi fa pcaK
Pectobacterium chrysanthe kdgR Pseudomonas putida pcaK
780
3226910
3227689
3347
6847 3227689 3226910 780 sp:KDGR_ERWCH RdgR

ABC transporter ATP-binding protein

305 547

63.6 57.2

32.5

Pseudomonas stutzeri

906 Sp:NOSF\_PSEST

gp:SCH10\_12

1584

3245342 3243759

3363 3362

25.2

Streptomyces coelicolor A3(2) SCH10.12

ABC transporter

5	on	36	mporter or d transporter2	permease	e component l		e component II	ferase	osphate d N-(5'- hranilate		beta chain	alpha chain	ane protein	nponent or enzyme II, A
10	Function	salicylate hydroxylase	proton/glutamate symporter or excitatory amino acid transporter2	tryptophan-specific permease	anthranilate synthase component I		anthranilate synthase component II	anthranilate phosphoribosyltransferase	indole-3-glycerol phosphate synthase (IGPS) and N-(5'- phosphoribosyl) anthranilate isomerase(PRAI)		tryptophan synthase beta chain	tryptophan synthase alpha chain	hypothetical membrane protein	PTS system, IIA component or unknown pentitol phosphotransferase enzyme II, A component
15	Matched length (a.a.)	476	507	170	515		208	348	474		417	283	521	152
20	Similarity (%)	49.4	54.4	99.4	9.66		100.0	99.4	98.3		97.9	96.5	86.8	71.7
	Identity (%)	28.2	25.4	99.4	99.2		99.0	99.4	97.3		97.6	95.4	9.99	30.3
os Table 1 (continued)	ns gene	ida	2	glutamicum	stofermentum		stofermentum	glutamicum	ctofermentum		ctofermentum	ctofermentum	licolor A3(2)	.12 ptxA
Table 1	Homologous gene	Pseudomonas putida	Homo sapiens eat2	Corynebacterium glutamicum AS019 ORF1	Brevibacterium lactofermentum trpE		Brevibacterium lactofermentum trpG	Corynebacterium glutamicum ATCC 21850 trpD	Brevibacterium lactofermentum trpC		Brevibacterium lactofermentum trpB	Brevibacterium lactofermentum trpA	Streptomyces coelicolor A3(2) SCJ21.17c	Escherichia coli K12 ptxA
<i>35</i>	db Match	prf:1706191A	SP:EAT2_HUMAN	pir.JC2326	Sp.TRPE_BRELA t		TRPG_BRELA	SP.TRPD_CORGL	SP:TRPC_BRELA		SP.TRPB_BRELA	SP:TRPA_BRELA	gp:SCJ21_17	sp:PTXA_ECOLI
	ORF (bp)	1326 prf.	1251 sp.	510 pir.	1554 sp.	171	624 TR	1044 sp.	1422 sp.	969	1251 sp.	840 sp	1539 gp	810 sp
45	Terminal (nt)	3230444	3231054 1	3233105	3234956 1	3233250	3235579	3236645 1	3238062	3236518	3239332 1	3240171	3240313 1	3241879
50	Initial (nt)	3229119	3232304	3232596	3233403	3233420	3234956	3235602	3236641	3237213	3238082	3239332	3241851	3242688
	SEQ NO.	+		6851	6852	6853	6854	6855	6856	6857	6858	6889	6860	6861
55	SEQ.	3349	3350	3351	3352	3353	3354	3355	3356	3357	3358	3359	3360	3361

5
10
15
20
25
30
35
40
45
50

	Function	cytchrome b6-F complex iron-sulfur subunit (Rieske iron-sulfur protein)	NADH oxidase or NADH-dependent flavin oxidoreductase	hypothetical membrane protein	hypothetical protein	bacterial regulatory protein, arsR family or methylenomycin A resistance protein	NADH oxidase or NADH-dependent flavin oxidoreductase	hypothetical protein					acetoin(diacetyl) reductase (acetoin dehydrogenase)	hypothetical protein	di-/tripeptide transpoter		bacterial regulatory protein, tetR family	hydroxyquinol 1,2-dioxygenase
	Matched length (a.a.)	305	336	328	262	102	347	226					238	85	469		188	246
	Similarity (%)	63.6	64.3	74.7	54.6	79.4	64.3	69.5					52.9	84.5	71.6		50.5	62.2
	Identity (%)	32.5	33.3	43.6	34.0	45.1	33.4	31.4					26.9	53.5	34.5		26.1	31.7
Table 1 (continued)	Homologous gene	Chlorobium limicola petC	Thermoanaerobacter brockii nadO	Escherichia coli K12 yfeH	Streptomyces coelicolor A3(2) SC111.36c	Streptomyces coelicolor Plasmid SCP1 mmr	Thermoanaerobacter brockii nadO	Saccharomyces cerevisiae ymyO					Klebsiella terrigena budC	Mycobacterium tuberculosis H37Rv Rv2094c	Lactococcus lactis subsp. lactis dtpT		Escherichia coli K12 acrR	Acinetobacter calcoaceticus catA
	db Match	sp:UCRI_CHLLT	sp:NADO_THEBR	Sp.YFEH_ECOLI	gp:SCI11_36	pir.A29606	sp:NADO_THEBR	sp:YMY0_YEAST		:			sp:BUDC_KLETE	sp:YY34_MYCTU	sp.DTPT_LACLA		sp:ACRR_ECOLI	sp:CATA_ACICA
	ORF (bp)	450	1110	972	774	348	1092	648	153	192	168	321	753	180	1359	171	555	903
	Terminal (nt)	3245766	3245822	3248205	3249165	3249187	3250742	3251405	3251466	3251743	3252133	3252316	3253480	3253739	3253824	3255719	3255744	3256471
	Initial (nt)	3245317	3246931	3247234	3248392	3249534	3249651	3250758	3251618	3251934	3252300	3252636	3252728	3253560	3255182	3255549	3256298	3257373
	SEQ NO. (a.a.)	6864	6865	6866	6867	6868	6989	6870	6871	6872	6873	6874	6875	6876	6877	6878	6819	6880
	SEQ NO (DNA)	3364	3365	3366	3367	3368	3369	3370	3371	3372	3373	3374	3375	3376	3377	3378	3379	3380

															· ·				
Function	maleyiacetate reductase	sugar transporter or D-xylose-proton symporter (D-xylose transporter)	bacterial transcriptional regulator or acetate operon repressor	oxidoreductase	diagnostic fragment protein sequence	myo-inositol 2-dehydrogenase	dehydrogenase or myo-inositol 2- dehydrogenase or streptomycin biosynthesis protein	phosphoesterase				stomatin		DEAD box RNA helicase family	hypothetical membrane protein		phosphomethylpyrimidine kinase	mercuric ion-binding protein or heavy-metal-associated domain containing protein	ectoine/proline uptake protein
Matched length (a.a.)	351	513	280	258	022	332	343	1242				206		1660	141		125	29	297
Similarity (%)	75.5	58.3	60.7	55.7	58.2	59.6	62.4	62.7				57.3		80.2	61.0		76.8	70.1	62.3
Identity (%)	43.0	31.4	25.7	27.2	25.9	26.5	34.1	33.3				28.6		58.4	34.8		50.4	46.3	29.9
Homologous gene	Pseudomonas sp. P51	Escherichia coli K12 xylE	Salmonella typhimurium iclR	Escherichia coli K12 ydgJ	Listeria innocua strain 4450	Sinorhizobium meliloti idhA	Streptomyces griseus strl	Bacillus subtilis yvnB				Caenorhabditis elegans unc1		Mycobacterium bovis BCG RvD1-Rv2024c	Mycobacterium leprae u2266k		Bacillus subtilis thiD	Bacillus subtilis yvgY	Corynebacterium glutamicum proP
db Match	sp:TCBF_PSESQ	sp:XYLE_ECOLI	sp:ICLR_SALTY	sp:YDGJ_ECOLI	gsp:W61761	sp:MI2D_BACSU	sp.STRI_STRGR	pir:C70044	*			Sp:UNC1_CAEEL		gp:MBO18605_3	prf:2323363AAM		sp:THID_BACSU	pir.F70041	prf.2501295A
ORF (bp)	1089	1524	861	1077	879	1005	1083	4032	645	618	1086	744	696	4929	507	360	909	243	837
Terminal (nt)	3257403	3258561	3261989	3263221	3264115	3265146	3266266	3271093	3267913	3268618	3272477	3274488	3275602	3276671	3281666	3283101	3282347	3283383	3283473
Initial (nt)	3258491	3260084	3261129	3262145	3263237	3264142	3265184	3267062	3268557	3269235	3271392	3275231	3276570	3281599	3282172	3282742	3282946	3283141	3284309
SEQ NO.	6881	6882	6883	6884	6885	6886	6887	6888	6889	6890	6891	6892	6893	6894	6895	9689	6897	6898	6899
SEQ NO.	3381	3382	3383	3384	3385	3386	3387	3388	3389	3390	3391	3392	3393	3394	3395	3396	3397	3398	3399

thioredoxin reductase

308

82.5

60.4

951 sp.TRXB\_STRCL Streptomyces clavuligerus trxB

5		Function	iron(III) dicitrate-binding periplasmic protein precursor or iron(III) dicitrate transport system permease protein	mitochondrial respiratory function protein or zinc-binding dehydrogenase or NADPH quinone oxidoreductase		-	phosphomethylpyrimidine kinase		mercuric ion-binding protein or heavy-metal-associated domain containing protein	branched-chain amino acid transport	branched-chain amino acid transport	rotein	IRNA nucleotidyltransferase	protein		hypothetical membrane protein	hypothetical membrane protein		RNA polymerase sigma-H factor or sigma-70 factor (ECF subfamily)
10			iron(III) dicitra protein precur transport syst	mitochondrial respirato protein or zinc-binding dehydrogenase or NAI oxidoreductase			phosphometh		mercuric ion-bindi heavy-metal-asso containing protein	branched-cha	branched-cha	hypothetical protein	tRNA nucleoti	mutator mutT protein		hypothetical n	hypothetical n		RNA polymer sigma-70 fact
15		Matched length (a.a.)	279	324			249		67	102	212	169	471	234		858	1201		189
20		Similarity (%)	9.09	58.0			75.5		70.1	65.7	0.79	56.2	51.8	69.2		54.3	60.1		6.09
		Identity (%)	29.4	27.2			46.2		41.8	36.3	32.1	23.7	26.8	43.6		25.8	35.7		30.2
25	Table 1 (continued)	Homologous gene	K12 fecB	nyces pombe			hiD		Ygv	azIO	aziD	K12 yqgE	K12 cca	uberculosis		uberculosis	uberculosis		eruginosa algU
30	Table 1	Homolog	Escherichia coli K12 fecB	Schizosaccharomyces pombe mrf1			Bacillus subtilis thiD		Bacillus subtilis yvgY	Bacillus subtilis azlD	Bacillus subtilis azID	Escherichia coli K12 yqgE	Escherichia coli K12 cca	Mycobacterium tuberculosis H37Rv Rv3908		Mycobacterium tuberculosis H37Rv Rv3909	Mycobacterium tuberculosis H37Rv Rv3910		Pseudomonas aeruginosa algU
<i>35</i>		db Match	sp:FECB_ECOL!	sp:MRF1_SCHPO			sp.THID_BACSU		pir.F70041	sp:AZLD_BACSU	sp:AZLC_BACSU	sp:YQGE_ECOLI	sp:ccA_EcoLI	pir.E70600		pir.F70600	pir.G70600		sp:RPSH_PSEAE
		ORF (bp)	957	1122	384	219	798	345	201	345	711	267	1320	996	273	2511	3249	723	603
45		Terminal (nt)	3284399	3286576	3287005	3287079	3287393	3288609	3288885	3288971	3289311	3290025	3290623	3293497	3292610	3296007	3299404	3298428	3300263
50		Initial (nt)	3285355	3285455	3286622	3287297	3288190	3288265	3288685	3289315	3290021	3290591	3291942	3292532	3292882	3293497	3296156	3297706	3299661
		SEQ NO.	<del></del>	6901	6902	6903	6904	6905	9069	6907	8069	6069	6910	6911	6912	6913	6914	6915	6916
55		SEQ NO. (DNA)	3400	3401	3402	3403	3404	3405	3406	3407	3408	3409	3410	3411	3412	3413	3414	3415	3416

5		Function		thioredoxin ch2, M-type	N-acetylmuramoyl-L-alanine amidase			hypothetical protein	hypothetical protein	partitioning or sporulation protein	glucose inhibited division protein B	hypothetical membrane protein	ribonuclease P protein component	50S ribosomal protein L34			L-aspartate-alpha-decarboxylase precursor	2-isopropylmalate synthase	hypothetical protein	aspartate-semialdehyde dehydrogenase	3-dehydroquinase
		70		thioredo	N-acetylr amidase			hypothe	hypothe	partition	glucose	hypothe	ribonuc	50S ribo			L-asparta precursor	2-isopro	hypothe	aspartat dehydro	3-dehyd
15		Matched length (a.a.)		119	196			212	367	272	153	313	123	47			136	616	88	344	149
20		Similarity (%)		76.5	75.4			58.5	60.5	78.0	64.7	75.4	59.4	93.6			100.0	100.0	100.0	100.0	100.0
		Identity (%)		42.0	51.0			34.4	37.6	65.0	36.0	44.7	26.8	83.0			100.0	100.0	100.0	100.0	100.0
25	Table 1 (continued)	ons gene		reinhardtii thi2	wlB			uberculosis	ıtida ygi2	uberculosis	<12 gidB	uberculosis	μρΑ	vium rpmH			glutamicum	glutamicum A	glutamicum lavum) ATCC	glutamicum	glutamicum
30	Table 1	Homologous gene		Chlamydomonas reinhardtii thi2	Bacillus subtilis cwlB			Mycobacterium tuberculosis H37Rv Rv3916c	Pseudomonas putida ygi2	Mycobacterium tuberculosis H37Rv parB	Escherichia coli K12 gidB	Mycobacterium tuberculosis H37Rv Rv3921c	Bacillus subtilis rnpA	Mycobacterium avium rpmH			Corynebacterium glutamicum panD	Corynebacterium glutamicum ATCC 13032 leuA	Corynebacterium glutamicum (Brevibacterium flavum) ATCC 13032 orfX	Corynebacterium glutamicum asd	Corynebacterium glutamicum ASO19 aroD
<i>35</i> <i>40</i>		db Match		Sp.THI2_CHLRE	sp:CWLB_BACSU			pir:D70851	sp:YGI2_PSEPU	sp.YGI1_PSEPU	sp.GIDB_ECOLI	pir.A70852	SP.RNPA_BACSU	gp:MAU19185_1			gp:AF116184_1	p:LEU1_CORGL	p:YLEU_CORGL	sp:DHAS_CORGL	gp:AF124518_1
		ORF (bp)	1185	372 sp	1242 sp	777	1041	618 pi	1152 sp	837 sp	s 699	951 pi	399 st	336 gr	294	222	408 gr	1848 sp	255 sp	1032 sp	447 gp
45		Terminal (nt)	3300119	3301729	3302996	3301989	3304475	3302999	3303636	3304835	3305864	3306682	3307971	3308412	3309321	3308822	147573	266154	268814	271691	446521
50		fnitial (nt)	3301303	3301358	3301755	3302765	3303435	3303616	3304787	3305671	3306532	3307632	3308369	3308747	3309028	3309043	147980	268001	269068	270660	446075
		SEQ NO. (a.a.)	6918	6919	6920	6921	6922	6923	6924	6925	6926	6927	6928	6359	6930	6931	6932	6933	6934	6935	9269
55		SEQ NO. (DNA)	3418	3419	3420	3421	3422	3423	3424	3425	3426	3427	3428	3429	3430	3431	3432	3433	3434	3435	3436
Ju																			·		

												·				
5	•	Function	elongation factor Tu	preprotein translocase secY subuit	isocitrate dehydrogenase (oxalosuccinatedecarboxylase)	acyl-CoA carboxylase or biotin- binding protein	citrate synthase	putative binding protein or peptidyl- prolyl cis-trans isomerase	glycine betaine transporter	hypothetical membrane protein	L-lysine permease	aromatic amino acid permease	hypothetical protein	succinyl diaminopimelate desuccinylase	proline transport system	arginyl-tRNA synthetase
		פ	elon	prep	isoci (oxa	acyl- bindi	citral	puta proly	glyci	hypo	L-lys	агоп	hypo	sacc	profi	argin
15		Matched length (a.a.)	396	440	738	591	437	118	595	426	501	463	316	369	524	550
20		Similarity (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		Identity (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
25	Table 1 (continued)	us gene	glutamicum	glutamicum avum) MJ233	glutamicum	glutamicum 3C	glutamicum	glutamicum	glutamicum	glutamicum	glutamicum	glutamicum	glutamicum	glutamicum	glutamicum	glutamicum 59 argS
30	Table 1 (	Homologous gene	Corynebacterium glutamicum ATCC 13059 tuf	Corynebacterium glutamicum (Brevibacterium flavum) MJ233 secY	Corynebacterium glutamicum ATCC 13032 icd	Corynebacterium glutamicum ATCC 13032 accBC	Corynebacterium glutamicum ATCC 13032 gltA	Corynebacterium glutamicum ATCC 13032 fkbA	Corynebacterium glutamicum ATCC 13032 betP	Corynebacterium glutamicum ATCC 13032 orf2	Corynebacterium glutamicum ATCC 13032 lysl	Corynebacterium glutamicum ATCC 13032 aroP	Corynebacterium glutamicum ATCC 13032 orf3	Corynebacterium glutamicum ATCC 13032 dapE	Corynebacterium glutamicum ATCC 13032 putP	Corynebacterium glutamicum AS019 ATCC 13059 argS
35 40		db Match	p.EFTU_CORGL	sp.SECY_CORGL	sp:IDH_CORGL	prf.2223173A	sp.CISY_CORGL	sp:FKBP_CORGL	p:BETP_CORGL	p:YLI2_CORGL	sp:LYSI_CORGL	sp:AROP_CORGL	pir.S52753	prf.2106301A	gp:CGPUTP_1	sp:SYR_CORGL
		ORF (bp)	1188 sı	1320 sı	2214 s <sub>l</sub>	1773 p	1311 sı	354 sı	1785 sı	1278 sı	1503 sı	1389 sı	948 pi	1107 pi	1572 gi	1650 sp
45		Terminal (nt)	527563	570771	677831	718580	879148	879629	946780	1029006	1030369	1153295	1154729	1156837	1218031	1239923
50		Initial (nt)	526376	569452	680044	720352	877838	879276	944996	1030283	1031871	1154683	1155676	1155731	1219602	1238274
		SEQ NO. (a.a.)	6937	6938	6633	6940	6941	6942	6943	6944	6945	6946	6947	6948	6949	0569
55		SEQ NO. (DNA)	3437	3438	3439	3440	3441	3442	3443	3444	3445	3446	3447	3448	3449	3450

	_															
5		Function	(DAP) neso- decarboxylase)	drogenase	e S	nit	rotein	ulator protein	acetohydroxy acid synthase, large subunit	acetohydroxy acid synthase, small subunit	acetohydroxy acid isomeroreductase	3-isopropylmalate dehydrogenase	PTS system, phosphoenolpyruvate sugar phosphotransferase (mannose and glucose transport)	dnase	oyltransferase	)r
10		P	diaminopimelate (DAP) decarboxylase (meso- diaminopimelate decarboxylase)	homoserine dehydrogenase	homoserine kinase	ion channel subunit	lysine exporter protein	lysine export regulator protein	acetohydroxy aci subunit	acetohydroxy aci subunit	acetohydroxy aci	3-isopropylmalat	PTS system, phosphoenolpyruva sugar phosphotransferase (mannose and glucose transport)	acetylglutamate kinase	ornithine carbamoyltransferase	arginine repressor
15		Matched length (a.a.)	445	445	. 608	216	236	290	626	172	338	340	683	294	319	171
20		Similarity (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		Identity (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
25	lable I (collillueu)	Homologous gene	m glutamicum 3059 lysA	m glutamicum 3059 hom	m glutamicum 3059 thrB	m glutamicum	m glutamicum	m glutamicum	m glutamicum /B	m glutamicum	m glutamicum /C	m glutamicum uB	m glutamicum M	m glutamicum gB	m glutamicum gF	m glutamicum
30 <u>4</u>	ומחום	olomoH	Corynebacterium glutamicum AS019 ATCC 13059 lysA	Corynebacterium glutamicum AS019 ATCC 13059 hom	Corynebacterium glutamicum AS019 ATCC 13059 thrB	Corynebacterium glutamicum R127 orf3	Corynebacterium glutamicum R127 lysE	Corynebacterium glutamicum R127 lysG	Corynebacterium ATCC 13032 ilvB	Corynebacterium glutamicum ATCC 13032 ilvN	Corynebacterium glutamicum ATCC 13032 ilvC	Corynebacterium glutamicum ATCC 13032 leuB	Corynebacterium glutamicum KCTC1445 ptsM	Corynebacterium glutamicum ATCC 13032 argB	Corynebacterium glutamicum ATCC 13032 argF	Corynebacterium glutamicum ASO19 argR
<i>35</i>		db Match	sp:DCDA_CORGL	sp.DHOM_CORGL	sp:KHSE_CORGL	gsp:W37716	sp.LYSE_CORGL	sp:LYSG_CORGL	sp:ILVB_CORGL	pir.B48648	pir.C48648	sp:LEU3_CORGL	prf:2014259A	sp:ARGB_CORGL	sp.OTCA_CORGL	gp:AF041436_1
		ORF (bp)	1335	1335	927	627	708	870	1878	516	1014	1020	2049	882	957	513
<b>4</b> 5		Terminal (nt)	1241263	1243841	1244781	1328243	1328246	1329884	1340008	1340540	1341737	1354508	1425265	1467372	1469521	1470040
50		Initial (nt)	1239929	1242507	1243855	1327617	1328953	1329015	1338131	1340025	1340724	1353489	1423217	1466491	1468565	1469528
		SEQ NO.	6951	6952	6953	6954	6955	9569	6957	6958	6369	0969	6961	6962	6963	6964
55		SEQ NO.	3451	3452	3453	3454	3455	3456	3457	3458	3459	3460	3461	3462	3463	3464

								<del></del>					$\overline{}$			
5		LC .	98	Se	poxylase	rotein, high	rane protein	e carboxylase	: (5- 3-phosphate	ase	polymerase	otein		ynthase	eductase	nase (acceptor)
10		Function	NADH dehydrogenase	phosphoribosyl-ATP- pyrophosphohydrolase	ornithine-cyclodecarboxylase	ammonium uptake protein, high affinity	protein-export membrane protein secG	phosphoenolpyruvate carboxylase	chorismate synthase (5- enolpyruvylshikimate-3-phosphate phospholyase)	restriction endonuclease	sigma factor or RNA transcription factor	glutamate-binding protein	recA protein	dihydrodipicolinate synthase	dihydrodipicolinate reductase	L-malate dehydrogenase (acceptor)
15		Matched length (a.a.)	467	87	362	452	77	919	410	632	331	295	376	301	248	200
20	ļ	Similarity (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		Identity (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
<i>25</i>	ntinued)	gene	utamicum	utamicum	utamicum	utamicum	utamicum	utamicum	utamicum	utamicum	utamicum	utamicum	utamicum	utamicum tofermentum)	utamicum tofermentum)	utamicum
30	Table 1 (continued)	Homologous gene	Corynebacterium glutamicum ATCC 13032 ndh	Corynebacterium glutamicum ASO19 hisE	Corynebacterium glutamicum ATCC 13032 ocd	Corynebacterium glutamicum ATCC 13032 amt	Corynebacterium glutamicum ATCC 13032 secG	Corynebacterium glutamicum ATCC 13032 ppc	Corynebacterium glutamicum AS019 aroC	Corynebacterium glutamicum ATCC 13032 cgllIR	Corynebacterium glutamicum ATCC 13869 sigB	Corynebacterium glutamicum ATCC 13032 gluB	Corynebacterium glutamicum AS019 recA	Corynebacterium glutamicum (Brevibacterium lactofermentum) ATCC 13869 dapA	Corynebacterium glutamicum (Brevibacterium lactofermentum) ATCC 13869 dapB	Corynebacterium glutamicum R127 mqo
35						<b></b>				l v						
40		db Match	gp:CGL238250_1	gp:AF086704_1	gp:CGL007732_4	gp:CGL007732_3	gp:CGL007732_2	prf: 1509267A	gp:AF124600_1	pir.855225	prf:2204286D	sp:GLUB_CORGL	sp:RECA_CORGL	sp:DAPA_BRELA	sp:DAPB_CORGL	gp:CGA224946_1
		ORF (bp)	1401	261	1086	1356	231	2757	1230	1896	993	885	1128	903	744	1500
45		Terminal (nt)	1543154	1586465	1674123	1675268	1677049	1677387	1719669	1882385	2021846	2061504	2063989	2079281	2081191	2113864
<i>50</i> .		Initial (nt)	1544554	1586725	1675208	1676623	1677279	1680143	1720898	1880490	2020854	2060620	2065116	2080183	2081934	2115363
		SEQ NO.	6965	9969	2969	6968	6969	6970	6971	6972	6973	6974	6975	6976	7269	6978
55		SEQ NO.	3465	3466	3467	3468	3469	3470	3471	3472	3473	3474	3475	3476	3477	3478

	_									····							
5		Function	rase, uridilylyl- me	nitrogen regulatory protein P-II	nsporter	glutamate dehydrogenase (NADP+)	Ð		hetase	hase	ectoine/proline/glycine betaine carrier	95		nase	cystathionine gamma-synthase	reductase	
10			uridilylyltransferase, removing enzyme	nitrogen regula	ammonium transporter	glutamate dehy	pyruvate kinase	glucokinase	glutamine synthetase	threonine synthase	ectoine/proline carrier	malate synthase	isocitrate lyase	glutamate 5-kinase	cystathionine g	ribonucleotide reductase	glutaredoxin
15		Matched length (a.a.)	692	112	438	447	475	323	477	481	615	739	432	369	386	148	77
20		Similarity (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		Identity (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
30 official 20	continued	us gene	glutamicum )	glutamicum 3	glutamicum P	glutamicum A	glutamicum	glutamicum	glutamicum A	glutamicum	glutamicum	glutamicum B	glutamicum A	glutamicum B	glutamicum	glutamicum	glutamicum H
30	ומחום ו	Homologous gene	Corynebacterium glutamicum ATCC 13032 glnD	Corynebacterium glutamicum ATCC 13032 glnB	Corynebacterium glutamicum ATCC 13032 amtP	Corynebacterium glutamicum ATCC 17965 gdhA	Corynebacterium glutamicum AS019 pyk	Corynebacterium ATCC 13032 glk	Corynebacterium glutamicum ATCC 13032 glnA	Corynebacterium glutamicum thrC	Corynebacterium glutamicum ATCC 13032 ectP	Corynebacterium glutamicum ATCC 13032 aceB	Corynebacterium glutamicum ATCC 13032 aceA	Corynebacterium glutamicum ATCC 17965 proB	Corynebacterium glutamicum ASO19 metB	Corynebacterium glutamicum ATCC 13032 nrdl	Corynebacterium glutamicum ATCC 13032 nrdH
<i>35</i>		db Match	gp:CAJ10319_4	gp:CAJ10319_3	gp:CAJ10319_2	pir.S32227	Sp.KPYK_CORGL	gp:AF096280_1	prf:2322244A	sp:THRC_CORGL	prf:2501295B	pir:140715	pir:140713	sp.PROB_CORGL	gp.AF126953_1	gp:AF112535_2	gp:AF112535_1
		ORF (bp)	2076 91	336 91	1314 g	1341 pi	1425 sı	696 696	1431 p	1443 s	1845 p	2217 p	1296 р	1107 S	1158 g	444 9	231 g
45		Terminal (nt)	2169666	2171751	2172154	2194742	2205668	2316582	2350259	2353600	2448328	2467925	2472035	2496670	2590312	2679684	2680419
50		Initial (nt)	2171741	2172086	2173467	2196082	2207092	2317550	2348829	2355042	2450172	2470141	2470740	2497776	2591469	2680127	2680649
		SEQ NO. (a.a.)	6269	6980	6981	6982	6983	6984	6985	9869	6987	6988	6869	0669	6991	6992	6993
55		SEQ NO.	3479	3480	3481	3482	3483	3484	3485	3486	3487	3488	3489	3490	3491	3492	3493
_ <del>_</del>		_		_													

	-
5	
5	
10	
15	
20	
25	
30	
35	
33	
40	
70	
45	
50	

_								<del></del>	
	Function	meso-diaminopimelate D- dehydrogenase	porin or cell wall channel forming protein	acetate kinase	phosphate acetyltransferase	multidrug resistance protein or macrolide-efflux pump or drug:proton antiporter	ATP-dependent protease regulatory subunit	prephenate dehydratase	ectoine/proline uptake protein
	Matched length (a.a.)	320	45	397	329	459	852	315	504
	Identity Similarity (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	Identity (%)	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Table 1 (continued)	Homologous gene	Corynebacterium glutamicum KY10755 ddh	Corynebacterium glutamicum MH20-22B porA	Corynebacterium glutamicum ATCC 13032 ackA	Corynebacterium glutamicum ATCC 13032 pta	Corynebacterium glutamicum ATCC 13032 cmr	Corynebacterium glutamicum ATCC 13032 clpB	Corynebacterium glutamicum pheA	Corynebacterium glutamicum ATCC 13032 proP
	db Match	sp:DDH_CORGL	gp:CGL238703_1	sp:ACKA_CORGL	prf.2516394A	prf.2309322A	sp:CLPB_CORGL	prf.1210266A	prf:2501295A
	ORF (bp)	960	135	1191	286	1377	2556	945	1512
	Terminal (nt)	2786756	2887944	2935315	2936508	2962718	2963606	3098578	3272563
	Initial (nt)	2787715	2888078	2936505	2937494	6998 2961342	2966161	7000 3099522	7001 3274074
	SEQ NO.	6994	6995	9669	2669	8669	6669	7000	7001
	SEQ NO (DNA)		3495	3496	3497	3498	3499	3500	3501

## Example 2

5

10

15

20

25

30

35

40

45

50

55

Determination of effective mutation site

(1) Identification of mutation site based on the comparison of the gene nucleotide sequence of lysine-producing B-6 strain with that of wild type strain ATCC 13032

[0374] Corynebacterium glutamicum B-6, which is resistant to S-(2-aminoethyl)cysteine (AEC), rifampicin, streptomycin and 6-azauracil, is a lysine-producing mutant having been mutated and bred by subjecting the wild type ATCC 13032 strain to multiple rounds of random mutagenesis with a mutagen, N-methyl-N' -nitro-N-nitrosoguanidine (NTG) and screening (Appl. Microbiol. Biotechnol., 32: 269-273 (1989)). First, the nucleotide sequences of genes derived from the B-6 strain and considered to relate to the lysine production were determined by a method similar to the above. The genes relating to the lysine production include lysE and lysG which are lysine-excreting genes; ddh, dapA, hom and IysC (encoding diaminopimelate dehydrogenase, dihydropicolinate synthase, homoserine dehydrogenase and aspartokinase, respectively) which are lysine-biosynthetic genes; and pyc and zwf (encoding pyruvate carboxylase and glucose-6-phosphate dehydrogenase, respectively) which are glucose-metabolizing genes. The nucleotide sequences of the genes derived from the production strain were compared with the corresponding nucleotide sequences of the ATCC 13032 strain genome represented by SEQ ID NOS:1 to 3501 and analyzed. As a result, mutation points were observed in many genes. For example, no mutation site was observed in lysE, lysG, ddh, dapA, and the like, whereas amino acid replacement mutations were found in hom, lysC, pyc, zwf, and the like. Among these mutation points, those which are considered to contribute to the production were extracted on the basis of known biochemical or genetic information. Among the mutation points thus extracted, a mutation, Val59Ala, in hom and a mutation, Pro458Ser, in pyc were evaluated whether or not the mutations were effective according to the following method.

(2) Evaluation of mutation, Val59Ala, in hom and mutation, Pro458Ser, in pyc

[0375] It is known that a mutation in hom inducing requirement or partial requirement for homoserine imparts lysine productivity to a wild type strain (*Amino Acid Fermentation*, ed. by Hiroshi Aida *et al.*, Japan Scientific Societies Press). However, the relationship between the mutation, Val59Ala, in *hom* and lysine production is not known. It can be examined whether or not the mutation, Val59Ala, in *hom* is an effective mutation by introducing the mutation to the wild type strain and examining the lysine productivity of the resulting strain. On the other hand, it can be examined whether or not the mutation, Pro458Ser, in *pyc* is effective by introducing this mutation into a lysine-producing strain which has a deregulated lysine-bioxynthetic pathway and is free from the *pyc* mutation, and comparing the lysine productivity of the resulting strain with the parent strain. As such a lysine-producing bacterium, No. 58 strain (FERM BP-7134) was selected (hereinafter referred to the "lysine-producing No. 58 strain" or the "No. 58 strain"). Based on the above, it was determined that the mutation, Val59Ala, in *hom* and the mutation, Pro458Ser, in *pyc* were introduced into the wild type strain of *Corynebacterium glutamicum* ATCC 13032 (hereinafter referred to as the "wild type ATCC 13032 strain" or the "ATCC 13032 strain") and the lysine-producing No. 58 strain, respectively, using the gene replacement method. A plasmid vector pCES30 for the gene replacement for the introduction was constructed by the following method.

[0376] A plasmid vector pCE53 having a kanamycin-resistant gene and being capable of autonomously replicating in Coryneform bacteria (*Mol. Gen. Genet., 196*: 175-178 (1984)) and a plasmid pMOB3 (ATCC 77282) containing a levansucrase gene (*sacB*) of *Bacillus subtilis* (*Molecular Microbiology, 6*: 1195-1204 (1992)) were each digested with *Pstl.* Then, after agarose gel electrophoresis, a pCE53 fragment and a 2.6 kb DNA fragment containing *sacB* were each extracted and purified using GENECLEAN Kit (manufactured by BIO 101). The pCE53 fragment and the 2.6 kb DNA fragment were ligated using Ligation Kit ver. 2 (manufactured by Takara Shuzo), introduced into the ATCC 13032 strain by the electroporation method (*FEMS Microbiology Letters,* 65: 299 (1989)), and cultured on BYG agar medium (medium prepared by adding 10 g of glucose, 20 g of peptone (manufactured by Kyokuto Pharmaceutical), 5 g of yeast extract (manufactured by Difco), and 16 g of Bactoagar (manufactured by Difco) to 1 liter of water, and adjusting its pH to 7.2) containing 25 μg/ml kanamycin at 30°C for 2 days to obtain a transformant acquiring kanamycin-resistance. As a result of digestion analysis with restriction enzymes, it was confirmed that a plasmid extracted from the resulting transformant by the alkali SDS method had a structure in which the 2.6 kb DNA fragment had been inserted into the *Pst*l site of pCE53. This plasmid was named pCES30.

[0377] Next, two genes having a mutation point, hom and pyc, were amplified by PCR, and inserted into pCES30 according to the TA cloning method (Bio Experiment Illustrated vol. 3, published by Shujunsha). Specifically, pCES30 was digested with BamHI (manufactured by Takara Shuzo), subjected to an agarose gel electrophoresis, and extracted and purified using GENECLEAN Kit (manufactured by BIO 101). The both ends of the resulting pCES30 fragment were blunted with DNA Blunting Kit (manufactured by Takara Shuzo) according to the attached protocol. The blunt-ended pCES30 fragment was concentrated by extraction with phenol/chloroform and precipitation with ethanol, and allowed

to react in the presence of Taq polymerase (manufactured by Roche Diagnostics) and dTTP at 70°C for 2 hours so that a nucleotide, thymine (T), was added to the 3'-end to prepare a T vector of pCES30.

[0378] Separately, chromosomal DNA was prepared from the lysine-producing B-6 strain according to the method of Saito et al. (*Biochem. Biophys. Acta, 72*: 619 (1963)). Using the chromosomal DNA as a template, PCR was carried out with Pfu turbo DNA polymelase (manufactured by Stratagene). In the mutated *hom* gene, the DNAs having the nucleotide sequences represented by SEQ ID NOS:7002 and 7003 were used as the primer set. In the mutated *pyc* gene, the DNAs having the nucleotide sequences represented by SEQ ID NOS:7004 and 7005 were used as the primer set. The resulting PCR product was subjected to agarose gel electrophoresis, and extracted and purified using GENE-GLEAN Kit (manufactured by BIO 101). Then, the PCR product was allowed to react in the presence of Taq polymerase (manufactured by Roche Diagnostics) and dATP at 72°C for 10 minutes so that a nucleotide, adenine (A), was added to the 3'-end.

10

15

20

25

30

35

40

45

50

[0379] The above pCES30 T vector fragment and the mutated *hom* gene (1.7 kb) or mutated *pyc* gene (3.6 kb) to which the nucleotide A had been added of the PCR product were concentrated by extraction with phenol/chloroform and precipitation with ethanol, and then ligated using Ligation Kit ver. 2. The ligation products were introduced into the ATCC 13032 strain according to the electroporation method, and cultured on BYG agar medium containing  $25 \,\mu g/ml$  kanamycin at 30°C for 2 days to obtain kanamycin-resistant transformants. Each of the resulting transformants was cultured overnight in BYG liquid medium containing  $25 \,\mu g/ml$  kanamycin, and a plasmid was extracted from the culturing solution medium according to the alkali SDS method. As a result of digestion analysis using restriction enzymes, it was confirmed that the plasmid had a structure in which the 1.7 kb or 3.6 kb DNA fragment had been inserted into pCES30. The plasmids thus constructed were named respectively pChom59 and pCpyc458.

[0380] The introduction of the mutations to the wild type ATCC 13032 strain and the lysine-producing No. 58 strain according to the gene replacement method was carried out according to the following method. Specifically, pChom59 and pCpyc458 were introduced to the ATCC 13032 strain and the No. 58 strain, respectively, and strains in which the plasmid is integrated into the chromosomal DNA by homologous recombination were selected using the method of lkeda et al. (Microbiology 144: 1863 (1998)). Then, the stains in which the second homologous recombination was carried out were selected by a selection method, making use of the fact that the Bacillus subtilis levansucrase encoded by pCES30 produced a suicidal substance (J. of Bacteriol., 174: 5462 (1992)). Among the selected strains, strains in which the wild type hom and pyc genes possessed by the ATCC 13032 strain and the No. 58 strain were replaced with the mutated hom and pyc genes, respectively, were isolated. The method is specifically explained below.

[0381] One strain was selected from the transformants containing the plasmid, pChom59 or pCpyc458, and the selected strain was cultured in BYG medium containing 20 µg/ml kanamycin, and pCG11 (Japanese Published Examined Patent Application No. 91827/94) was introduced thereinto by the electroporation method. pCG11 is a plasmid vector having a spectinomycin-resistant gene and a replication origin which is the same as pCE53. After introduction of the pCGII, the strain was cultured on BYG agar medium containing 20 µg/ml kanamycin and 100 µg/ml spectinomycin at 30°C for 2 days to obtain both the kanamycin- and spectinomycin-resistant transformant. The chromosome of one strain of these transformants was examined by the Southern blotting hybridization according to the method reported by Ikeda *et al.* (*Microbiology, 144*: 1863 (1998)). As a result, it was confirmed that pChom59 or pCpyc458 had been integrated into the chromosome by the homologous recombination of the Cambell type. In such a strain, the wild type and mutated *hom* or *pyc* genes are present closely on the chromosome, and the second homologous recombination is liable to arise therebetween.

[0382] Each of these transformants (having been recombined once) was spread on Suc agar medium (medium prepared by adding 100 g of sucrose, 7 g of meat extract, 10 g of peptone, 3 g of sodium chloride, 5 g of yeast extract (manufactured by Difco), and 18 g of Bactoagar (manufactured by Difco) to 1 liter of water, and adjusting its pH 7.2) and cultured at 30°C for a day. Then the colonies thus growing were selected in each case. Since a strain in which the sacB gene is present converts sucrose into a suicide substrate, it cannot grow in this medium (J. Bacteriol., 174: 5462 (1992)). On the other hand, a strain in which the sacB gene was deleted due to the second homologous recombination between the wild type and the mutated hom or pyc genes positioned closely to each other forms no suicide substrate and, therefore, can grow in this medium. In the homologous recombination, either the wild type gene or the mutated gene is deleted together with the sacB gene. When the wild type is deleted together with the sacB gene, the gene replacement into the mutated type arises.

[0383] Chromosomal DNA of each the thus obtained second recombinants was prepared by the above method of Saito *et al.* PCR was carried out using Pfu turbo DNA polymerase (manufactured by Stratagene) and the attached buffer. In the *hom* gene, DNAs having the nucleotide sequences represented by SEQ ID NOS:7002 and 7003 were used as the primer set. Also, in the *pyc* gene was used, DNAs having the nucleotide sequences represented by SEQ ID NOS:7004 and 7005 were used as the primer set. The nucleotide sequences of the PCR products were determined by the conventional method so that it was judged whether the *hom* or *pyc* gene of the second recombinant was a wild type or a mutant. As a result, the second recombinant which were called HD-1 and No. 58pyc were target strains having the mutated *hom* gene and *pyc* gene, respectively.

(3) Lysine production test of HD-1 and No. 58pyc strains

[0384] The HD-1 strain (strain obtained by incorporating the mutation, Val59Ala, in the hom gene into the ATCC 13032 strain) and the No. 58pyc strain (strain obtained by incorporating the mutation, Pro458Ser, in the pyc gene into the lysine-producing No. 58 strain) were subjected to a culture test in a 5 I jar fermenter by using the ATCC 13032 strain and the lysine-producing No. 58 strain respectively as a control. Thus lysine production was examined. [0385] After culturing on BYG agar medium at 30°C for 24 hours, each strain was inoculated into 250 ml of a seed medium (medium prepared by adding 50 g of sucrose, 40 g of corn steep liquor, 8.3 g of ammonium sulfate, 1 g of urea, 2 g of potassium dihydrogenphosphate, 0.83 g of magnesium sulfate heptahydrate, 10 mg of iron sulfate heptahydrate, 1 mg of copper sulfate pentahydrate, 10 mg of zinc sulfate heptahydrate, 10 mg of  $\beta$ -alanine, 5 mg of nicotinic acid, 1.5 mg of thiamin hydrochloride, and 0.5 mg of biotin to 1 liter of water, and adjusting its pH to 7.2, then to which 30 g of calcium carbonate had been added) contained in a 2 1 buffle-attached Erlenmeyer flask and cultured therein at 30°C for 12 to 16 hours. A total amount of the seed culturing medium was inoculated into 1,400 ml of a main culture medium (medium prepared by adding 60 g of glucose, 20 g of corn steep liquor, 25 g of ammonium chloride, 2.5 g of potassium dihydrogenphosphate, 0.75 g of magnesium sulfate heptahydrate, 50 mg of iron sulfate heptahydrate, 13 mg of manganese sulfate pentahydrate, 50 mg of calcium chloride, 6.3 mg of copper sulfate pentahydrate, 1.3 mg of zinc sulfate heptahydrate, 5 mg of nickel chloride hexahydrate, 1.3 mg of cobalt chloride hexahydrate, 1.3 mg of ammonium molybdenate tetrahydrate, 14 mg of nicotinic acid, 23 mg of β-alanine, 7 mg of thiamin hydrochloride, and 0.42 mg of biotin to 1 liter of water) contained in a 5 1 jar fermenter and cultured therein at 32°C, 1 vvm and 800 rpm while controlling the pH to 7.0 with aqueous ammonia. When glucose in the medium had been consumed, a glucose feeding solution (medium prepared by adding 400 g glucose and 45 g of ammonium chloride to 1 liter of water) was continuously added. The addition of feeding solution was carried out at a controlled speed so as to maintain the dissolved oxygen concentration within a range of 0.5 to 3 ppm. After culturing for 29 hours, the culture was terminated. The cells were separated from the culture medium by centrifugation and then L-lysine hydrochloride in the supernatant was quantified by high performance liquid chromatography (HPLC). The results are shown in Table 2 below.

Table 2

Strain	L-Lysine hydrochloride yield (g/l)
ATCC 13032	0
HD-1	8
No. 58	45
No. 58pyc	51

[0386] As is apparent from the results shown in Table 2, the lysine productivity was improved by introducing the mutation, Val59Ala, in the *hom* gene or the mutation, Pro458Ser, in the pyc gene. Accordingly, it was found that the mutations are both effective mutations relating to the production of lysine. Strain, AHP-3, in which the mutation, Val59Ala, in the *hom* gene and the mutation, Pro458Ser, in the *pyc* gene have been introduced into the wild type ATCC 13032 strain together with the mutation, Thr331lle in the *lysC* gene has been deposited on December 5, 2000, in National Institute of Bioscience and Human Technology, Agency of Industrial Science and Technology (Higashi 1-1-3, Tsukuba-shi, Ibaraki, Japan) as FERM BP-7382.

#### Example 3

5

10

15

20

25

30

35

40

50

55

<sup>45</sup> Reconstruction of lysine-producing strain based on genome information

[0387] The lysine-producing mutant B-6 strain (*Appl. Microbiol. Biotechnol., 32*: 269-273 (1989)), which has been constructed by multiple round random mutagenesis with NTG and screening from the wild type ATCC 13032 strain, produces a remarkably large amount of lysine hydrochloride when cultured in a jar at 32°C using glucose as a carbon source. However, since the fermentation period is long, the production rate is less than 2.1 g/l/h. Breeding to reconstitute only effective mutations relating to the production of lysine among the estimated at least 300 mutations introduced into the B-6 strain in the wild type ATCC 13032 strain was performed.

(1) Identification of mutation point and effective mutation by comparing the gene nucleotide sequence of the B-6 strain with that of the ATCC 13032 strain

[0388] As described above, the nucleotide sequences of genes derived from the B-6 strain were compared with the

corresponding nucleotide sequences of the ATCC 13032 strain genome represented by SEQ ID NOS:1 to 3501 and analyzed to identify many mutation points accumulated in the chromosome of the B-6 strain. Among these, a mutation, Val591Ala, in *hom*, a mutation, Thr311lle, in *lysC*, a mutation, Pro458Ser, in *pyc* and a mutation, Ala213Thr, in *zwf* were specified as effective mutations relating to the production of lysine. Breeding to reconstitute the 4 mutations in the wild type strain and for constructing of an industrially important lysine-producing strain was carried out according to the method shown below.

(2) Construction of plasmid for gene replacement having mutated gene

5

15

20

25

30

45

- 10 [0389] The plasmid for gene replacement, pChom59, having the mutated *hom* gene and the plasmid for gene replacement, pCpyc458, having the mutated *pyc* gene were prepared in the above Example 2(2). Plasmids for gene replacement having the mutated *lysC* and *zwf* were produced as described below.
  - [0390] The *lysC* and *zwf* having mutation points were amplified by PCR, and inserted into a plasmid for gene replacement, pCES30, according to the TA cloning method described in Example 2(2) (Bio Experiment Illustrated, Vol. 3). [0391] Separately, chromosomal DNA was prepared from the lysine-producing B-6 strain according to the above method of Saito *et al.* Using the chromosomal DNA as a template, PCR was carried out with Pfu turbo DNA polymerase (manufactured by Stratagene). In the mutated *lysC* gene, the DNAs having the nucleotide sequences represented by SEQ ID NOS:7006 and 7007 were used as the primer set. In the mutated *zwf* gene, the DNAs having the nucleotide sequences represented by SEQ ID NOS:7008 and 7009 as the primer set. The resulting PCR product was subjected to agarose gel electrophoresis, and extracted and purified using GENEGLEAN Kit (manufactured by BIO 101). Then, the PCR product was allowed to react in the presence of Taq DNA polymerase (manufactured by Roche Diagnostics) and dATP at 72°C for 10 minutes so that a nucleotide, adenine (A), was added to the 3'-end.
  - [0392] The above pCES30 T vector fragment and the mutated *lysC* gene (1.5 kb) or mutated *zwf* gene (2.3 kb) to which the nucleotide A had been added of the PCR product were concentrated by extraction with phenol/chloroform and precipitation with ethanol, and then ligated using Ligation Kit ver. 2. The ligation products were introduced into the ATCC 13032 strain according to the electroporation method, and cultured on BYG agar medium containing 25 μg/ml kanamycin at 30°C for 2 days to obtain kanamycin-resistant transformants. Each of the resulting transformants was cultured overnight in BYG liquid medium containing 25 μg/ml kanamycin, and a plasmid was extracted from the culturing solution medium according to the alkali SDS method. As a result of digestion analysis using restriction enzymes, it was confirmed that the plasmid had a structure in which the 1.5 kb or 2.3 kb DNA fragment had been inserted into pCES30. The plasmids thus constructed were named respectively pClysC311 and pCzwf213.
  - (3) Introduction of mutation, Thr311lle, in IysC into one point mutant HD-1
- [0393] Since the one mutation point mutant HD-1 in which the mutation, Val59Ala, in hom was introduced into the wild type ATCC 13032 strain had been obtained in Example 2(2), the mutation, Thr311lle, in lysC was introduced into the HD-1 strain using pClysC311 produced in the above (2) according to the gene replacement method described in Example 2(2). PCR was carried out using chromosomal DNA of the resulting strain and, as the primer set, DNAs having the nucleotide sequences represented by SEQ ID NOS:7006 and 7007 in the same manner as in Example 2(2). As a result of the fact that the nucleotide sequence of the PCR product was determined in the usual manner, it was confirmed that the strain which was named AHD-2 was a two point mutant having the mutated lysC gene in addition to the mutated hom gene.
  - (4) Introduction of mutation, Pro458Ser, in pyc into two point mutant AHD-2
  - [0394] The mutation, Pro458Ser, in *pyc* was introduced into the AHD-2 strain using the pCpyc458 produced in Example 2(2) by the gene replacement method described in Example 2(2). PCR was carried out using chromosomal DNA of the resulting strain and, as the primer set, DNAs having the nucleotide sequences represented by SEQ ID NOS:7004 and 7005 in the same manner as in Example 2(2). As a result of the fact that the nucleotide sequence of the PCR product was determined in the usual manner, it was confirmed that the strain which was named AHD-3 was a three point mutant having the mutated *pyc* gene in addition to the mutated *hom* gene and *lysC* gene.
  - (5) Introduction of mutation, Ala213Thr, in zwf into three point mutant AHP-3
- [0395] The mutation, Ala213Thr, in zwf was introduced into the AHP-3 strain using the pCzwf458 produced in the above (2) by the gene replacement method described in Example 2(2). PCR was carried out using chromosomal DNA of the resulting strain and, as the primer set, DNAs having the nucleotide sequences represented by SEQ ID NOS: 7008 and 7009 in the same manner as in Example 2(2). As a result of the fact that the nucleotide sequence of the PCR

product was determined in the usual manner, it was confirmed that the strain which was named APZ-4 was a four point mutant having the mutated *zwf* gene in addition to the mutated *hom* gene, *lysC* gene and *pyc* gene.

(6) Lysine production test on HD-1, AHD-2, AHP-3 and APZ-4 strains

[0396] The HD-1, AHD-2, AHP-3 and APZ-4 strains obtained above were subjected to a culture test in a 5 I jar fermenter in accordance with the method of Example 2(3).

[0397] Table 3 shows the results.

5

10

15

20

25

30

35

40

45

50

55

Table 3

Strain	L-Lysine hydrochloride (g/l)	Productivity (g/l/h)
HD-1	8	0.3
AHD-2	73	2.5
AHP-3	80	2.8
APZ-4	86	3.0

[0398] Since the lysine-producing mutant B-6 strain which has been bred based on the random mutation and selection shows a productivity of less than 2.1 g/l/h, the APZ-4 strain showing a high productivity of 3.0 g/l/h is useful in industry.

(7) Lysine fermentation by APZ-4 strain at high temperature

[0399] The APZ-4 strain, which had been reconstructed by introducing 4 effective mutations into the wild type strain, was subjected to the culturing test in a 5 I jar fermenter in the same manner as in Example 2(3), except that the culturing temperature was changed to  $40^{\circ}$ C.

[0400] The results are shown in Table 4.

Table 4

Temperature (°C)	L-Lysine hydrochloride (g/l)	Productivity (g/l/h)
32	86	3.0
40	95	3.3

[0401] As is apparent from the results shown in Table 4, the lysine hydrochloride titer and productivity in culturing at a high temperature of 40°C comparable to those at 32°C were obtained. In the mutated and bred lysine-producing B-6 strain constructed by repeating random mutation and selection, the growth and the lysine productivity are lowered at temperatures exceeding 34°C so that lysine fermentation cannot be carried out, whereas lysine fermentation can be carried out using the APZ-4 strain at a high temperature of 40°C so that the load of cooling is greatly reduced and it is industrially useful. The lysine fermentation at high temperatures can be achieved by reflecting the high temperature adaptability inherently possessed by the wild type strain on the APZ-4 strain.

[0402] As demonstrated in the reconstruction of the lysine-producing strain, the present invention provides a novel breeding method effective for eliminating the problems in the conventional mutants and acquiring industrially advantageous strains. This methodology which reconstitutes the production strain by reconstituting the effective mutation is an approach which is efficiently carried out using the nucleotide sequence information of the genome disclosed in the present invention, and its effectiveness was found for the first time in the present invention.

# Example 4

Production of DNA microarray and use thereof

**[0403]** A DNA microarray was produced based on the nucleotide sequence information of the ORF deduced from the full nucleotide sequences of *Corynebacterium glutamicum* ATCC 13032 using software, and genes of which expression is fluctuated depending on the carbon source during culturing were searched.

(1) Production of DNA microarray

[0404] Chromosomal DNA was prepared from Corynebacterium glutamicum ATCC 13032 by the method of Saito et

al. (Biochem. Biophys. Acta, 72: 619 (1963)). Based on 24 genes having the nucleotide sequences represented by SEQ ID NOS:207, 3433, 281, 3435, 3439, 765, 3445, 1226, 1229, 3448, 3451, 3453, 3455, 1743, 3470, 2132, 3476, 3477, 3485, 3488, 3489, 3494, 3496, and 3497 from the ORFs shown in Table 1 deduced from the full genome nucleotide sequence of Corynebacterium glutamicum ATCC 13032 using software and the nucleotide sequence of rabbit globin gene (GenBank Accession No. V00882) used as an internal standard, oligo DNA primers for PCR amplification represented by SEQ ID NOS:7010 to 7059 targeting the nucleotide sequences of the genes were synthesized in a usual manner.

[0405] As the oligo DNA primers used for the PCR,

10

20

30

35

50

55

[0406] DNAs having the nucleotide sequence represented by SEQ ID NOS:7010 and 7011 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:207,

[0407] DNAs having the nucleotide sequence represented by SEQ ID NOS:7012 and 7013 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3433,

[0408] DNAs having the nucleotide sequence represented by SEQ ID NOS:7014 and 7015 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:281,

[0409] DNAs having the nucleotide sequence represented by SEQ ID NOS:7016 and 7017 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3435,

[0410] DNAs having the nucleotide sequence represented by SEQ ID NOS:7018 and 7019 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3439,

[0411] DNAs having the nucleotide sequence represented by SEQ ID NOS:7020 and 7021 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:765,

[0412] DNAs having the nucleotide sequence represented by SEQ ID NOS:7022 and 7023 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3445,

[0413] DNAs having the nucleotide sequence represented by SEQ ID NOS:7024 and 7025 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:1226,

[0414] DNAs having the nucleotide sequence represented by SEQ ID NOS:7026 and 7027 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:1229,

[0415] DNAs having the nucleotide sequence represented by SEQ ID NOS:7028 and 7029 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3448,

[0416] DNAs having the nucleotide sequence represented by SEQ ID NOS:7030 and 7031 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3451,

[0417] DNAs having the nucleotide sequence represented by SEQ ID NOS:7032 and 7033 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3453,

[0418] DNAs having the nucleotide sequence represented by SEQ ID NOS:7034 and 7035 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3455,

[0419] DNAs having the nucleotide sequence represented by SEQ ID NOS:7036 and 7037 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:1743,

[0420] DNAs having the nucleotide sequence represented by SEQ ID NOS:7038 and 7039 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3470,

[0421] DNAs having the nucleotide sequence represented by SEQ ID NOS:7040 and 7041 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:2132,

[0422] DNAs having the nucleotide sequence represented by SEQ ID NOS:7042 and 7043 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3476,

[0423] DNAs having the nucleotide sequence represented by SEQ ID NOS:7044 and 7045 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3477,

[0424] DNAs having the nucleotide sequence represented by SEQ ID NOS:7046 and 7047 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3485,

[0425] DNAs having the nucleotide sequence represented by SEQ ID NOS:7048 and 7049 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3488,

[0426] DNAs having the nucleotide sequence represented by SEQ ID NOS:7050 and 7051 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3489,

[0427] DNAs having the nucleotide sequence represented by SEQ ID NOS:7052 and 7053 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3494,

[0428] DNAs having the nucleotide sequence represented by SEQ ID NOS:7054 and 7055 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3496,

[0429] DNAs having the nucleotide sequence represented by SEQ ID NOS:7056 and 7057 were used for the amplification of the DNA having the nucleotide sequence represented by SEQ ID NO:3497, and

[0430] DNAs having the nucleotide sequence represented by SEQ ID NOS:7058 and 7059 were used for the amplification of the DNA having the nucleotide sequence of the rabbit globin gene,

as the respective primer set.

[0431] The PCR was carried for 30 cycles with each cycle consisting of 15 seconds at 95°C and 3 minutes at 68°C using a thermal cycler (GeneAmp PCR system 9600, manufactured by Perkin Elmer), TaKaRa EX-Taq (manufactured by Takara Shuzo), 100 ng of the chromosomal DNA and the buffer attached to the TaKaRa Ex-Taq reagent. In the case of the rabbit globin gene, a single-stranded cDNA which had been synthesized from rabbit globin mRNA (manufactured by Life Technologies) according to the manufacture's instructions using a reverse transcriptase RAV-2 (manufactured by Takara Shuzo). The PCR product of each gene thus amplified was subjected to agarose gel electrophoresis and extracted and purified using QIAquick Gel Extraction Kit (manufactured by QIAGEN). The purified PCR product was concentrated by precipitating it with ethanol and adjusted to a concentration of 200 ng/µl. Each PCR product was spotted on a slide glass plate (manufactured by Matsunami Glass) having MAS coating in 2 runs using GTMASS SYSTEM (manufactured by Nippon Laser & Electronics Lab.) according to the manufacture's instructions.

# (2) Synthesis of fluorescence labeled cDNA

[0432] The ATCC 13032 strain was spread on BY agar medium (medium prepared by adding 20 g of peptone (manufactured by Kyokuto Pharmaceutical), 5 g of yeast extract (manufactured by Difco), and 16 g of Bactoagar (manufactured by Difco) to in 1 liter of water and adjusting its pH to 7.2) and cultured at 30°C for 2 days. Then, the cultured strain was further inoculated into 5 ml of BY liquid medium and cultured at 30°C overnight. Then, the cultured strain was further inoculated into 30 ml of a minimum medium (medium prepared by adding 5 g of ammonium sulfate, 5 g of urea, 0.5 g of monopotassium dihydrogenphosphate, 0.5 g of dipotassium monohydrogenphosphate, 20.9 g of morpholinopropanesulfonic acid, 0.25 g of magnesium sulfate heptahydrate, 10 mg of calcium chloride dihydrate, 10 mg of manganese sulfate monohydrate, 10 mg of ferrous sulfate heptahydrate, 1 mg of zinc sulfate heptahydrate, 0.2 mg copper sulfate, and 0.2 mg biotin to 1 liter of water, and adjusting its pH to 6.5) containing 110 mmol/l glucose or 200 mmol/l ammonium acetate, and cultured in an Erlenmyer flask at 30° to give 1.0 of absorbance at 660 nm. After the cells were prepared by centrifuging at 4°C and 5,000 rpm for 10 minutes, total RNA was prepared from the resulting cells according to the method of Bormann et al. ( Molecular Microbiology, 6: 317-326 (1992)). To avoid contamination with DNA, the RNA was treated with Dnasel (manufactured by Takara Shuzo) at 37°C for 30 minutes and then further purified using Qiagen RNeasy MiniKit (manufactured by QIAGEN) according to the manufacture's instructions. To 30 μg of the resulting total RNA, 0.6 μl of rabbit globin mRNA (50 ng/μl, manufactured by Life Technologies) and 1 μl of a random 6 mer primer (500 ng/μl, manufactured by Takara Shuzo) were added for denaturing at 65°C for 10 minutes, followed by quenching on ice. To the resulting solution, 6 µl of a buffer attached to Superscript II (manufactured by Lifetechnologies), 3 μl of 0.1 mol/l DTT, 1.5 μl of dNTPs (25 mmol/l dATP, 25 mmol/l dCTP, 25 mmol/l dGTP, 10 mmol/ I dTTP), 1.5 μI of Cy5-dUTP or Cy3-dUTP (manufactured by NEN) and 2 μI of Superscript II were added, and allowed to stand at 25°C for 10 minutes and then at 42°C for 110 minutes. The RNA extracted from the cells using glucose as the carbon source and the RNA extracted from the cells using ammonium acetate were labeled with Cy5-dUTP and Cy3-dUTP, respectively. After the fluorescence labeling reaction, the RNA was digested by adding 1.5 μl of 1 mol/l sodium hydroxide-20 mmol/l EDTA solution and 3.0 µl of 10% SDS solution, and allowed to stand at 65°C for 10 minutes. The two cDNA solutions after the labeling were mixed and purified using Qiagen PCR purification Kit (manufactured by QIAGEN) according to the manufacture's instructions to give a volume of 10 μl.

## (3) Hybridization

[0433] UltraHyb (110  $\mu$ l) (manufactured by Ambion) and the fluorescence-labeled cDNA solution (10  $\mu$ l) were mixed and subjected to hybridization and the subsequent washing of slide glass using GeneTAC Hybridization Station (manufactured by Genomic Solutions) according to the manufacture's instructions. The hybridization was carried out at 50°C, and the washing was carried out at 25°C.

# (4) Fluorescence analysis

50 **[0434]** The fluorescence amount of each DNA array having the fluorescent cDNA hybridized therewith was measured using ScanArray 4000 (manufactured by GSI Lumonics).

[0435] Table 5 shows the Cy3 and Cy5 signal intensities of the genes having been corrected on the basis of the data of the rabbit globin used as the internal standard and the Cy3/Cy5 ratios.

Table 5

SEQ ID NO	Cy3 intensity	Cy5 intensity	Cy3/Cy5
207	5248	3240	1.62

10

15

20

25

35

40

Table 5 (continued)

SEQ ID NO	Cy3 intensity	Cy5 intensity	Cy3/Cy5
3433	2239	2694	0.83
281	2370	2595	0.91
3435	2566	2515	1.02
3439	5597	6944	0.81
765	6134	4943	1.24
3455	1169	1284	0.91
1226	1301	1493	0.87
1229	1168	1131	1.03
3448	1187	1594	0.74
3451	2845	3859	0.74
3453	3498	1705	2.05
3455	1491	1144	1.30
1743	1972	1841	1.07
3470	4752	3764	1.26
2132	1173	1085	1.08
3476	1847	1420	1.30
3477	1284	1164	1.10
3485	4539	8014	0.57
3488	34289	1398	24.52
3489	43645	1497	29.16
3494	3199	2503	1.28
3496	3428	2364	1.45
3497	3848	3358	1.15

[0436] The ORF function data estimated by using software were searched for SEQ ID NOS:3488 and 3489 showing remarkably strong Cy3 signals. As a result, it was found that SEQ ID NOS:3488 and 3489 are a maleate synthase gene and an isocitrate lyase gene, respectively. It is known that these genes are transcriptionally induced by acetic acid in *Corynebacterium glutamicum* (*Archives of Microbiology, 168*: 262-269 (1997)).

[0437] As described above, a gene of which expression is fluctuates could be discovered by synthesizing appropriate oligo DNA primers based on the ORF nucleotide sequence information deduced from the full genomic nucleotide sequence information of *Corynebacterium glutamicum* ATCC 13032 using software, amplifying the nucleotide sequences of the gene using the genome DNA of *Corynebacterium glutamicum* as a template in the PCR reaction, and thus producing and using a DNA microarray.

[0438] This Example shows that the expression amount can be analyzed using a DNA microarray in the 24 genes. On the other hand, the present DNA microarray techniques make it possible to prepare DNA microarrays having thereon several thousand gene probes at once. Accordingly, it is also possible to prepare DNA microarrays having thereon all of the ORF gene probes deduced from the full genomic nucleotide sequence of *Corynebacterium glutamicum* ATCC 13032 determined by the present invention, and analyze the expression profile at the total gene level of *Corynebacterium glutamicum* using these arrays.

# Example 5

10

15

20

25

30

35

45

50

55

Homology search using Corynebacterium glutamicum genome sequence

# (1) Search of adenosine deaminase

[0439] The amino acid sequence (ADD\_ECOLI) of *Escherichia coli* adenosine deaminase was obtained from Swissprot Database as the amino acid sequence of the protein of which function had been confirmed as adenosine deaminase (EC3.5.4.4). By using the full length of this amino acid sequence as a query, a homology search was carried out on a nucleotide sequence database of the genome sequence of *Corynebacterium glutamicum* or a database of the amino acids in the ORF region deduced from the genome sequence using FASTA program (*Proc. Natl. Acad. Sci. ISA, 85*: 2444-2448 (1988)). A case where E-value was le<sup>-10</sup> or less was judged as being significantly homologous. As a result,

no sequence significantly homologous with the *Escherichia coli* adenosine deaminase was found in the nucleotide sequence database of the genome sequence of *Corynebacterium glutamicum* or the database of the amino acid sequences in the ORF region deduced from the genome sequence. Based on these results, it is assumed that *Corynebacterium glutamicum* contains no ORF having adenosine deaminase activity and thus has no activity of converting adenosine into inosine.

## (2) Search of glycine cleavage enzyme

5

10

15

20

25

30

35

40

45

55

**[0440]** The sequences (GCSP\_ECOLI, GCST\_ECOLI and GCSH\_ECOLI) of glycine decarboxylase, aminomethyl transferase and an aminomethyl group carrier each of which is a component of *Escherichia coli* glycine cleavage enzyme as the amino acid sequence of the protein, of which function had been confirmed as glycine cleavage enzyme (EC2.1.2.10), were obtained from Swiss-prot Database.

[0441] By using these full-length amino acid sequences as a query, a homology search was carried out on a nucleotide sequence database of the genome sequence of *Corynebacterium glutamicum* or a database of the ORF amino acid sequences deduced from the genome sequence using FASTA program. A case where E-value was le<sup>-10</sup> or less was judged as being significantly homologous. As a result, no sequence significantly homologous with the glycine decarboxylase, the aminomethyl transferase or the aminomethyl group carrier each of which is a component of *Escherichia coli* glycine cleavage enzyme, was found in the nucleotide sequence database of the genome sequence of *Corynebacterium glutamicum* or the database of the ORF amino acid sequences estimated from the genome sequence. Based on these results, it is assumed that *Corynebacterium glutamicum* contains no ORF having the activity of glycine decarboxylase, aminomethyl transferase or the aminomethyl group carrier and thus has no activity of the glycine cleavage enzyme.

## (3) Search of IMP dehydrogenase

[0442] The amino acid sequence (IMDH ECOLI) of Escherichia coli IMP dehydrogenase as the amino acid sequence of the protein, of which function had been confirmed as IMP dehydrogenase (EC1.1.1.205), was obtained from Swissprot Database. By using the full length of this amino acid sequence as a query, a homology search was carried out on a nucleotide sequence database of the genome sequence of Corynebacterium glutamicum or a database of the ORF amino acid sequences predicted from the genome sequence using FASTA program. A case where E-value was le-10 or less was judged as being significantly homologous. As a result, the amino acid sequences encoded by two ORFs, namely, an ORF positioned in the region of the nucleotide sequence No. 615336 to 616853 (or ORF having the nucleotide sequence represented by SEQ ID NO:672) and another ORF positioned in the region of the nucleotide sequence No. 616973 to 618094 (or ORF having the nucleotide sequence represented by SEQ ID NO:674) were significantly homologous with the ORFs of Escherichia coli IMP dehydrogenase. By using the above-described predicted amino acid sequence as a query in order to examine the similarity of the amino acid sequences encoded by the ORFs with IMP dehydrogenases of other organisms in greater detail, a search was carried out on GenBank (http://www.ncbi.nlm. nih.gov/) nr-aa database (amino acid sequence database constructed on the basis of GenBankCDS translation products, PDB database, Swiss-Prot database, PIR database, PRF database by eliminating duplicated registrations) using BLAST program. As a result, both of the two amino acid sequences showed significant homologies with IMP dehdyrogenases of other organisms and clearly higher homologies with IMP dehdyrogenases than with amino acid sequences of other proteins, and thus, it was assumed that the two ORFs would function as IMP dehydrogenase. Based on these results, it was therefore assumed that Corynebacterium glutamicum has two ORFs having the IMP dehydrogenase activity.

# Example 6

Proteome analysis of proteins derived from Corynebacterium glutamicum

50 (1) Preparations of proteins derived from Corynebacterium glutamicum ATCC 13032, FERM BP-7134 and FERM BP-158

[0443] Culturing tests of Corynebacterium glutamicum ATCC 13032 (wild type strain), Corynebacterium glutamicum FERM BP-7134 (lysine-producing strain) and Corynebacterium glutamicum (FERM BP-158, lysine-highly producing strain) were carried out in a 5 l jar fermenter according to the method in Example 2(3). The results are shown in Table 6.

Table 6

Strain	L-Lysine yield (g/l)	
ATCC 13032	0	
FERM BP-7134	45	
FERM BP-158	60	

[0444] After culturing, cells of each strain were recovered by centrifugation. These cells were washed with Tris-HCl buffer (10 mmol/l Tris-HCl, pH 6.5, 1.6 mg/ml protease inhibitor (COMPLETE; manufactured by Boehringer Mannheim)) three times to give washed cells which could be stored under freezing at -80°C. The freeze-stored cells were thawed before use, and used as washed cells.

[0445] The washed cells described above were suspended in a disruption buffer (10 mmol/l Tris-HCl, pH 7.4, 5 mmol/l magnesium chloride, 50 mg/l RNase, 1.6 mg/ml protease inhibitor (COMPLETE: manufactured by Boehringer Mannheim)), and disrupted with a disruptor (manufactured by Brown) under cooling. To the resulting disruption solution, DNase was added to give a concentration of 50 mg/l, and allowed to stand on ice for 10 minutes. The solution was centrifuged (5,000  $\times$  g, 15 minutes, 4°C) to remove the undisrupted cells as the precipitate, and the supernatant was recovered.

[0446] To the supernatant, urea was added to give a concentration of 9 mol/l, and an equivalent amount of a lysis buffer (9.5 mol/l urea, 2% NP-40, 2% Ampholine, 5% mercaptoethanol, 1.6 mg/ml protease inhibitor (COMPLETE; manufactured by Boehringer Mannheim) was added thereto, followed by thoroughly stirring at room temperature for dissolving.

[0447] After being dissolved, the solution was centrifuged at  $12,000 \times g$  for 15 minutes, and the supernatant was recovered.

[0448] To the supernatant, ammonium sulfate was added to the extent of 80% saturation, followed by thoroughly stirring for dissolving.

**[0449]** After being dissolved, the solution was centrifuged (16,000  $\times$  g, 20 minutes, 4°C), and the precipitate was recovered. This precipitate was dissolved in the lysis buffer again and used in the subsequent procedures as a protein sample. The protein concentration of this sample was determined by the method for quantifying protein of Bradford.

(2) Separation of protein by two dimensional electrophoresis

5

10

15

20

25

30

35

40

45

50

[0450] The first dimensional electrophoresis was carried out as described below by the isoelectric electrophoresis method.

[0451] A molded dry IPG strip gel (pH 4-7, 13 cm, Immobiline DryStrips; manufactured by Amersham Pharmacia Biotech) was set in an electrophoretic apparatus (Multiphor II or IPGphor; manufactured by Amersham Pharmacia Biotech) and a swelling solution (8 mol/l urea, 0.5% Triton X-100, 0.6% dithiothreitol, 0.5% Ampholine, pH 3-10) was packed therein, and the gel was allowed to stand for swelling 12 to 16 hours.

[0452] The protein sample prepared above was dissolved in a sample solution (9 mol/l urea, 2% CHAPS, 1% dithiothreitol, 2% Ampholine, pH 3-10), and then about 100 to 500  $\mu$ g (in terms of protein) portions thereof were taken and added to the swollen IPG strip gel.

[0453] The electrophoresis was carried out in the 4 steps as defined below under controlling the temperature to 20°C:

step 1: 1 hour under a gradient mode of 0 to 500V;

step 2: 1 hour under a gradient mode of 500 to 1,000 V;

step 3: 4 hours under a gradient mode of 1,000 to 8,000 V; and

step 4: 1 hour at a constant voltage of 8,000 V.

[0454] After the isoelectric electrophoresis, the IPG strip gel was put off from the holder and soaked in an equilibration buffer A (50 mmol/l Tris-HCl, pH 6.8, 30% glycerol, 1% SDS, 0.25% dithiothreitol) for 15 minutes and another equilibration buffer B (50 mmol/l Tris-HCl, pH 6.8, 6 mol/l urea, 30% glycerol, 1% SDS, 0.45% iodo acetamide) for 15 minutes to sufficiently equilibrate the gel.

[0455] After the equilibrium, the IPG strip gel was lightly rinsed in an SDS electrophoresis buffer (1.4% glycine, 0.1% SDS, 0.3% Tris-HCl, pH 8.5), and the second dimensional electrophoresis depending on molecular weight was carried out as described below to separate the proteins.

[0456] Specifically, the above IPG strip gel was closely placed on 14% polyacrylamide slub gel (14% polyacrylamide, 0.37% bisacrylamide, 37.5 mmol/l Tris-HCl, pH 8.8, 0.1% SDS, 0.1% TEMED, 0.1% ammonium persulfate) and sub-

jected to electrophoresis under a constant voltage of 30 mA at 20°C for 3 hours to separate the proteins.

(3) Detection of protein spot

15

20

25

30

35

40

45

- [0457] Coomassie staining was performed by the method of Gorg et al. (*Electrophoresis, 9*: 531-546 (1988)) for the slub gel after the second dimensional electrophoresis. Specifically, the slub gel was stained under shaking at 25°C for about 3 hours, the excessive coloration was removed with a decoloring solution, and the gel was thoroughly washed with distilled water.
- [0458] The results are shown in Fig. 2. The proteins derived from the ATCC 13032 strain (Fig. 2A), FERM BP-7134 strain (Fig. 2B) and FERM BP-158 strain (Fig. 2C) could be separated and detected as spots.
  - (4) In-gel digestion of detected protein spot
  - [0459] The detected spots were each cut out from the gel and transferred into siliconized tube, and 400 μl of 100 mmol/1 ammonium bicarbonate: acetonitrile solution (1:1, v/v) was added thereto, followed by shaking overnight and freeze-dried as such. To the dried gel, 10 μl of a lysylendopeptidase (LysC) solution (manufactured by WAKO, prepared with 0.1% SDS-containing 50 mmol/l ammonium bicarbonate to give a concentration of 100 ng/μl) was added and the gel was allowed to stand for swelling at 0°C for 45 minutes, and then allowed to stand at 37°C for 16 hours. After removing the LysC solution, 20 μl of an extracting solution (a mixture of 60% acetonitrile and 5% formic acid) was added, followed by ultrasonication at room temperature for 5 minutes to disrupt the gel. After the disruption, the extract was recovered by centrifugation (12,000 rpm, 5 minutes, room temperature). This operation was repeated twice to recover the whole extract. The recovered extract was concentrated by centrifugation *in vacuo* to halve the liquid volume. To the concentrate, 20 μl of 0.1% trifluoroacetic acid was added, followed by thoroughly stirring, and the mixture was subjected to desalting using ZipTip (manufactured by Millipore). The protein absorbed on the carriers of ZipTip was eluted with 5 μl of α-cyano-4-hydroxycinnamic acid for use as a sample solution for analysis.
  - (5) Mass spectrometry and amino acid sequence analysis of protein spot with matrix assisted laser desorption ionization time of flight mass spectrometer (MALDI-TOFMS)
  - [0460] The sample solution for analysis was mixed in the equivalent amount with a solution of a peptide mixture for mass calibration (300 nmol/l Angiotensin II, 300 nmol/l Neurotensin, 150 nmol/l ACTHclip 18-39, 2.3 μmol/l bovine insulin B chain), and 1 μl of the obtained solution was spotted on a stainless probe and crystallized by spontaneously drying.
    - [0461] As measurement instruments, REFLEX MALDI-TOF mass spectrometer (manufactured by Bruker) and an N2 laser (337 nm) were used in combination.
    - [0462] The analysis by PMF (peptide-mass finger printing) was carried out using integration spectra data obtained by measuring 30 times at an accelerated voltage of 19.0 kV and a detector voltage of 1.50 kV under reflector mode conditions. Mass calibration was carried out by the internal standard method.
    - [0463] The PSD (post-source decay) analysis was carried out using integration spectra obtained by successively altering the reflection voltage and the detector voltage at an accelerated voltage of 27.5 kV.
    - [0464] The masses and amino acid sequences of the peptide fragments derived from the protein spot after digestion were thus determined.
    - (6) Identification of protein spot
    - [0465] From the amino acid sequence information of the digested peptide fragments derived from the protein spot obtained in the above (5), ORFs corresponding to the protein were searched on the genome sequence database of *Corynebacterium glutamicum* ATCC 13032 as constructed in Example 1 to identify the protein.
- [0466] The identification of the protein was carried out using MS-Fit program and MS-Tag program of intranet protein prospector.
  - (a) Search and identification of gene encoding high-expression protein
  - [0467] In the proteins derived from *Corynebacterium glutamicum* ATCC 13032 showing high expression amounts in CBB-staining shown in Fig. 2A, the proteins corresponding to Spots-1, 2, 3, 4 and 5 were identified by the above method. [0468] As a result, it was found that Spot-1 corresponded to enolase which was a protein having the amino acid sequence of SEQ ID NO:4585; Spot-2 corresponded to phosphoglycelate kinase which was a protein having the amino acid sequence of SEQ ID NO:5254; Spot-3 corresponded to glyceraldehyde-3-phosphate dehydrogenase which was

a protein having the amino acid sequence represented by SEQ ID NO:5255; Spot-4 corresponded to fructose bisphosphate aldolase which was a protein having the amino acid sequence represented by SEQ ID NO:6543; and Spot-5 corresponded to triose phosphate isomerase which was a protein having the amino acid sequence represented by SEQ ID NO:5252.

- [0469] These genes, represented by SEQ ID NOS:1085, 1754, 1775, 3043 and 1752 encoding the proteins corresponding to Spots-1, 2, 3, 4 and 5, respectively, encoding the known proteins are important in the central metabolic pathway for maintaining the life of the microorganism. Particularly, it is suggested that the genes of Spots-2, 3 and 5 form an operon and a high-expression promoter is encoded in the upstream thereof (*J. of Eacteriol., 174*: 6067-6086 (1992)).
- [0470] Also, the protein corresponding to Spot-9 in Fig. 2 was identified in the same manner as described above, and it was found that Spot-9 was an elongation factor Tu which was a protein having the amino acid sequence represented by SEQ ID No:6937, and that the protein was encoded by DNA having the nucleotide sequence represented by SEQ ID No:3437.
- [0471] Based on these results, the proteins having high expression level were identified by proteome analysis using the genome sequence database of *Corynebacterium glutamicum* constructed in Example 1. Thus, the nucleotide sequences of the genes encoding the proteins and the nucleotide sequences upstream thereof could be searched simultaneously. Accordingly, it is shown that nucleotide sequences having a function as a high-expression promoter can be efficiently selected.
- 20 (b) Search and identification of modified protein
  - [0472] Among the proteins derived from *Corynebacterium glutamicum* FERM BP-7134 shown in Fig. 2B, Spots-6, 7 and 8 were identified by the above method. As a result, these three spots all corresponded to catalase which was a protein having the amino acid sequence represented by SEQ ID NO:3785.
- [0473] Accordingly, all of Spots-6, 7 and 8 detected as spots differing in isoelectric mobility were all products derived from a catalase gene having the nucleotide sequence represented by SEQ ID No:285. Accordingly, it is shown that the catalase derived from *Corynebacterium glutamicum* FERM BP-7134 was modified after the translation.
  - **[0474]** Based on these results, it is confirmed that various modified proteins can be efficiently searched by proteome analysis using the genome sequence database of *Corynebacterium glutamicum* constructed in Example 1.
  - (c) Search and identification of expressed protein effective in lysine production
  - [0475] It was found out that in Fig. 2A (ATCC 13032: wild type strain), Fig. 2B (FERM BP-7134: lysine-producing strain) and Fig. 2C (FERM BP-158: lysine-highly producing strain), the catalase corresponding to Spot-8 and the elongation factor Tu corresponding to Spot-9 as identified above showed the higher expression level with an increase in the lysine productivity.
    - **[0476]** Based on these results, it was found that hopeful mutated proteins can be efficiently searched and identified in breeding aiming at strengthening the productivity of a target product by the proteome analysis using the genome sequence database of *Corynebacterium glutamicum* constructed in Example 1.
- 40 [0477] Moreover, useful mutation points of useful mutants can be easily specified by searching the nucleotide sequences (nucleotide sequences of promoter, ORF, or the like) relating to the identified proteins using the above database and using primers designed on the basis of the sequences. As a result of the fact that the mutation points are specified, industrially useful mutants which have the useful mutations or other useful mutations derived therefrom can be easily bred.
- [0478] While the invention has been described in detail and with reference to specific embodiments thereof, it will be apparent to one of skill in the art that various changes and modifications can be made therein without departing from the spirit and scope thereof. All references cited herein are incorporated in their entirety.

# 50 Claims

55

30

- 1. A method for at least one of the following:
  - (A) identifying a mutation point of a gene derived from a mutant of a coryneform bacterium,
  - (B) measuring an expression amount of a gene derived from a coryneform bacterium,
  - (C) analyzing an expression profile of a gene derived from a coryneform bacterium,
  - (D) analyzing expression patterns of genes derived from a coryneform bacterium, or
  - (E) identifying a gene homologous to a gene derived from a coryneform bacterium,

# said method comprising:

5

10

15

20

25

30

35

45

55

- (a) producing a polynucleotide array by adhering to a solid support at least two polynucleotides selected from the group consisting of first polynucleotides comprising the nucleotide sequence represented by any one of SEQ ID NOS:1 to 3501, second polynucleotides which hybridize with the first polynucleotides under stringent conditions, and third polynucleotides comprising a sequence of 10 to 200 continuous bases of the first or second polynucleotides,
- (b) incubating the polynucleotide array with at least one of a labeled polynucleotide derived from a coryneform bacterium, a labeled polynucleotide derived from a mutant of the coryneform bacterium or a labeled polynucleotide to be examined, under hybridization conditions,
- (c) detecting any hybridization, and
- (d) analyzing the result of the hybridization.
- 2. The method according to claim 1, wherein the coryneform bacterium is a microorganism belonging to the genus Corynebacterium, the genus Brevibacterium, or the genus Microbacterium.
  - 3. The method according to claim 2, wherein the microorganism belonging to the genus Corynebacterium is selected from the group consisting of Corynebacterium glutamicum, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum, Corynebacterium callunae, Corynebacterium herculis, Corynebacterium lilium, Corynebacterium melassecola, Corynebacterium thermoaminogenes, and Corynebacterium ammoniagenes.
  - 4. The method according to claim 1, wherein the polynucleotide derived from a coryneform bacterium, the polynucleotide derived from a mutant of the coryneform bacterium or the polynucleotide to be examined is a gene relating to the biosynthesis of at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogues thereof.
  - 5. The method according to claim 1, wherein the polynucleotide to be examined is derived from Escherichia coli.
  - 6. A polynucleotide array, comprising:

at least two polynucleotides selected from the group consisting of first polynucleotides comprising the nucleotide sequence represented by any one of SEQ ID NOS:1 to 3501, second polynucleotides which hybridize with the first polynucleotides under stringent conditions, and third polynucleotides comprising 10 to 200 continuous bases of the first or second polynucleotides, and a solid support adhered thereto.

- 7. A polynucleotide comprising the nucleotide sequence represented by SEQ ID NO:1 or a polynucleotide having a homology of at least 80% with the polynucleotide.
- 8. A polynucleotide comprising any one of the nucleotide sequences represented by SEQ ID NOS:2 to 3431, or a polynucleotide which hybridizes with the polynucleotide under stringent conditions.
  - A polynucleotide encoding a polypeptide having any one of the amino acid sequences represented by SEQ ID NOS:3502 to 6931, or a polynucleotide which hybridizes therewith under stringent conditions.
  - 10. A polynucleotide which is present in the 5' upstream or 3' downstream of a polynucleotide comprising the nucleotide sequence of any one of SEQ ID NOS:2 to 3431 in a whole polynucleotide comprising the nucleotide sequence represented by SEQ ID NO:1, and has an activity of regulating an expression of the polynucleotide.
- 11. A polynucleotide comprising 10 to 200 continuous bases in the nucleotide sequence of the polynucleotide of any one of claims 7 to 10, or a polynucleotide comprising a nucleotide sequence complementary to the polynucleotide comprising 10 to 200 continuous based.
  - 12. A recombinant DNA comprising the polynucleotide of any one of claims 8 to 11.
  - 13. A transformant comprising the polynucleotide of any one of claims 8 to 11 or the recombinant DNA of claim 12.
  - 14. A method for producing a polypeptide, comprising:

culturing the transformant of claim 13 in a medium to produce and accumulate a polypeptide encoded by the polynucleotide of claim 8 or 9 in the medium, and recovering the polypeptide from the medium.

15. A method for producing at least one of an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogues thereof, comprising:

culturing the transformant of claim 13 in a medium to produce and accumulate at least one of an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogues thereof in the medium, and recovering the at least one of the amino acid, the nucleic acid, the vitamin, the saccharide, the organic acid, and analogues thereof from the medium.

- 16. A polypeptide encoded by a polynucleotide comprising the nucleotide sequence selected from SEQ ID NOS:2 to 3431.
- 17. A polypeptide comprising the amino acid sequence selected from SEQ ID NOS:3502 to 6931.
- 18. The polypeptide according to claim 16 or 17, wherein at least one amino acid is deleted, replaced, inserted or added, said polypeptides having an activity which is substantially the same as that of the polypeptide without said at least one amino acid deletion, replacement, insertion or addition.
- 19. A polypeptide comprising an amino acid sequence having a homology of at least 60% with the amino acid sequence of the polypeptide of claim 16 or 17, and having an activity which is substantially the same as that of the polypeptide.
- 25 20. An antibody which recognizes the polypeptide of any one of claims 16 to 19.
  - 21. A polypeptide array, comprising:

10

15

20

30

35

45

50

55

at least one polypeptide or partial fragment polypeptide selected from the polypeptides of claims 16 to 19 and partial fragment polypeptides of the polypeptides, and a solid support adhered thereto.

22. A polypeptide array, comprising:

at least one antibody which recognizes a polypeptide or partial fragment polypeptide selected from the polypeptides of claims 16 to 19 and partial fragment polypeptides of the polypeptides, and a solid support adhered thereto.

- 23. A system based on a computer for identifying a target sequence or a target structure motif derived from a coryneform bacterium, comprising the following:
  - (i) a user input device that inputs at least one nucleotide sequence information selected from SEQ ID NOS:1 to 3501, and target sequence or target structure motif information;
  - (ii) a data storage device for at least temporarily storing the input information;
  - (iii) a comparator that compares the at least one nucleotide sequence information selected from SEQ ID NOS: 1 to 3501 with the target sequence or target structure motif information, recorded by the data storage device for screening and analyzing nucleotide sequence information which is coincident with or analogous to the target sequence or target structure motif information; and
  - (iv) an output device that shows a screening or analyzing result obtained by the comparator.
  - 24. A method based on a computer for identifying a target sequence or a target structure motif derived from a coryneform bacterium, comprising the following:
    - (i) inputting at least one nucleotide sequence information selected from SEQ ID NOS:1 to 3501, target sequence information or target structure motif information into a user input device;
    - (ii) at least temporarily storing said information;
    - (iii) comparing the at least one nucleotide sequence information selected from SEQ ID NOS:1 to 3501 with the target sequence or target structure motif information; and

- (iv) screening and analyzing nucleotide sequence information which is coincident with or analogous to the target sequence or target structure motif information.
- 25. A system based on a computer for identifying a target sequence or a target structure motif derived from a coryneform bacterium, comprising the following:
  - (i) a user input device that inputs at least one amino acid sequence information selected from SEQ ID NOS: 3502 to 7001, and target sequence or target structure motif information;
  - (ii) a data storage device for at least temporarily storing the input information;
  - (iii) a comparator that compares the at least one amino acid sequence information selected from SEQ ID NOS: 3502 to 7001 with the target sequence or target structure motif information, recorded by the data storage device for screening and analyzing amino acid sequence information which is coincident with or analogous to the target sequence or target structure motif information; and
  - (iv) an output device that shows a screening or analyzing result obtained by the comparator.
  - 26. A method based on a computer for identifying a target sequence or a target structure motif derived from a coryneform bacterium, comprising the following:
    - (i) inputting at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001, and target sequence information or target structure motif information into a user input device;
    - (ii) at least temporarily storing said information;

5

10

15

20

25

30

35

45

50

- (iii) comparing the at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001 with the target sequence or target structure motif information; and
- (iv) screening and analyzing amino acid sequence information which is coincident with or analogous to the target sequence or target structure motif information.
- 27. A system based on a computer for determining a function of a polypeptide encoded by a polynucleotide having a target nucleotide sequence derived from a coryneform bacterium, comprising the following:
  - (i) a user input device that inputs at least one nucleotide sequence information selected from SEQ ID NOS:2 to 3501, function information of a polypeptide encoded by the nucleotide sequence, and target nucleotide sequence information;
  - (ii) a data storage device for at least temporarily storing the input information;
  - (iii) a comparator that compares the at least one nucleotide sequence information selected from SEQ ID NOS: 2 to 3501 with the target nucleotide sequence information for determining a function of a polypeptide encoded by a polynucleotide having the target nucleotide sequence which is coincident with or analogous to the polynucleotide having at least one nucleotide sequence selected from SEQ ID NOS:2 to 3501; and
  - (iv) an output devices that shows a function obtained by the comparator.
- 28. A method based on a computer for determining a function of a polypeptide encoded by a polypeptide encoded by a polypucleotide having a target nucleotide sequence derived from a coryneform bacterium, comprising the following:
  - (i) inputting at least one nucleotide sequence information selected from SEQ ID NOS:2 to 3501, function information of a polypeptide encoded by the nucleotide sequence, and target nucleotide sequence information;
  - (ii) at least temporarily storing said information;
  - (iii) comparing the at least one nucleotide sequence information selected from SEQ ID NOS:2 to 3501 with the target nucleotide sequence information; and
  - (iv) determining a function of a polypeptide encoded by a polynucleotide having the target nucleotide sequence which is coincident with or analogous to the polynucleotide having at least one nucleotide sequence selected from SEQ ID NOS:2 to 3501.
  - 29. A system based on a computer for determining a function of a polypeptide having a target amino acid sequence derived from a coryneform bacterium, comprising the following:
    - (i) a user input device that inputs at least one amino acid sequence information selected from SEQ ID NOS: 3502 to 7001, function information based on the amino acid sequence, and target amino acid sequence information:

- (ii) a data storing device for at least temporarily storing the input information;
- (iii) a comparator that compares the at least one amino acid sequence information selected from SEQ ID NOS: 3502 to 7001 with the target amino acid sequence information for determining a function of a polypeptide having the target amino acid sequence which is coincident with or analogous to the polypeptide having at least one amino acid sequence selected from SEQ ID NOS:3502 to 7001; and
- (iv) an output device that shows a function obtained by the comparator.
- **30.** A method based on a computer for determining a function of a polypeptide having a target amino acid sequence derived from a coryneform bacterium, comprising the following:
  - (i) inputting at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001, function information based on the amino acid sequence, and target amino acid sequence information;
  - (ii) at least temporarily storing said information;

5

10

15

25

30

35

- (iii) comparing the at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001 with the target amino acid sequence information; and
- (iv) determining a function of a polypeptide having the target amino acid sequence which is coincident with or analogous to the polypeptide having at least one amino acid sequence selected from SEQ ID NOS:3502 to 7001.
- 20 31. The system according to any one of claims 23, 25, 27 and 29, wherein a coryneform bacterium is a microorganism of the genus *Corynebacterium*, the genus *Brevibacterium*, or the genus *Microbacterium*.
  - **32.** The method according to any one of claims 24, 26, 28 and 30, wherein a coryneform bacterium is a microorganism of the genus *Corynebacterium*, the genus *Brevibacterium*, or the genus *Microbacterium*.
  - 33. The system according to claim 31, wherein the microorganism belonging to the genus Corynebacterium is selected from the group consisting of Corynebacterium glutamicum, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum, Corynebacterium callunae, Corynebacterium herculis, Corynebacterium lilium, Corynebacterium melassecola, Corynebacterium thermoaminogenes, and Corynebacterium ammoniagenes.
  - 34. The method according to claim 32, wherein the microorganism belonging to the genus Corynebacterium is selected from the group consisting of Corynebacterium glutamicum, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum, Corynebacterium callunae, Corynebacterium herculis, Corynebacterium lilium, Corynebacterium melassecola, Corynebacterium thermoaminogenes, and Corynebacterium ammoniagenes.
  - 35. A recording medium or storage device which is readable by a computer in which at least one nucleotide sequence information selected from SEQ ID NOS:1 to 3501 or function information based on the nucleotide sequence is recorded, and is usable in the system of claim 23 or 27 or the method of claim 24 or 28.
- 36. A recording medium or storage device which is readable by a computer in which at least one amino acid sequence information selected from SEQ ID NOS:3502 to 7001 or function information based on the amino acid sequence is recorded, and is usable in the system of claim 25 or 29 or the method of claim 26 or 30.
- 37. The recording medium or storage device according to claim 35 or 36, which is a computer readable recording medium selected from the group consisting of a floppy disc, a hard disc, a magnetic tape, a random access memory (RAM), a read only memory (ROM), a magneto-optic disc (MO), CD-ROM, CD-R, CD-RW, DVD-ROM, DVD-RAM and DVD-RW.
- 38. A polypeptide having a homoserine dehydrogenase activity, comprising an amino acid sequence in which the Val residue at the 59th in the amino acid sequence of homoserine dehydrogenase derived from a coryneform bacterium is replaced with an amino acid residue other than a Val residue.
  - 39. A polypeptide comprising an amino acid sequence in which the Val residue at the 59th position in the amino acid sequence as represented by SEQ ID NO:6952 is replaced with an amino acid residue other than a Val residue.
  - **40.** The polypeptide according to claim 38 or 39, wherein the Val residue at the 59th position is replaced with an Ala residue.

- 41. A polypeptide having pyruvate carboxylase activity, comprising an amino acid sequence in which the Pro residue at the 458th position in the amino acid sequence of pyruvate carboxylase derived from a coryneform bacterium is replaced with an amino acid residue other than a Pro residue.
- 42. A polypeptide comprising an amino acid sequence in which the Pro residue at the 458th position in the amino acid sequence represented by SEQ ID NO:4265 is replaced with an amino acid residue other than a Pro residue.
  - 43. The polypeptide according to claim 41 or 42, wherein the Pro residue at the 458th position is replaced with a Ser residue.
  - 44. The polypeptide according to any one of claims 38 to 43, which is derived from Corynebacterium glutamicum.
  - 45. A DNA encoding the polypeptide of any one of claims 38 to 44.
- 46. A recombinant DNA comprising the DNA of claim 45.

10

20

30

35

40

45

55

- 47. A transformant comprising the recombinant DNA of claim 46.
- 48. A transformant comprising in its chromosome the DNA of claim 45.
- 49. The transformant according to claim 47 or 48, which is derived from a coryneform bacterium.
- 50. The transformant according to claim 49, which is derived from Corynebacterium glutamicum.
- 25 51. A method for producing L-lysine, comprising:

culturing the transformant of any one of claims 47 to 50 in a medium to produce and accumulate L-lysine in the medium, and recovering the L-lysine from the culture.

- **52.** A method for breeding a coryneform bacterium using the nucleotide sequence information represented by SEQ ID NOS:1 to 3431, comprising the following:
  - (i) comparing a nucleotide sequence of a genome or gene of a production strain derived a coryneform bacterium which has been subjected to mutation breeding so as to produce at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogous thereof by a fermentation method, with a corresponding nucleotide sequence in SEQ ID NOS:1 to 3431;
  - (ii) identifying a mutation point present in the production strain based on a result obtained by (i);
  - (iii) introducing the mutation point into a coryneform bacterium which is free of the mutation point, or deleting the mutation point from a coryneform bacterium having the mutation point; and
  - (iv) examining productivity by the fermentation method of the compound selected in (i) of the coryneform bacterium obtained in (iii).
- 53. The method according to claim 52, wherein the gene is a gene encoding an enzyme in a biosynthetic pathway or a signal transmission pathway.
  - 54. The method according to claim 52, wherein the mutation point is a mutation point relating to a useful mutation which improves or stabilizes the productivity.
- 55. A method for breading a coryneform bacterium using the nucleotide sequence information represented by SEQ ID NOS:1 to 3431, comprising:
  - (i) comparing a nucleotide sequence of a genome or gene of a production strain derived a coryneform bacterium which has been subjected to mutation breeding so as to produce at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogous thereof by a fermentation method, with a corresponding nucleotide sequence in SEQ ID NOS:1 to 3431;
  - (ii) identifying a mutation point present in the production strain based on a result obtain by (i);
  - (iii) deleting a mutation point from a coryneform bacterium having the mutation point; and

- (iv) examining productivity by the fermentation method of the compound selected in (i) of the coryneform bacterium obtained in (iii).
- **56.** The method according to claim 55, wherein the gene is a gene encoding an enzyme in a biosynthetic pathway or a signal transmission pathway.
- 57. The method according to claim 55, wherein the mutation point is a mutation point which decreases or destabilizes the productivity.
- 58. A method for breeding a coryneform bacterium using the nucleotide sequence information represented by SEQ ID NOS:2 to 3431, comprising the following:
  - (i) identifying an isozyme relating to biosynthesis of at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogous thereof, based on the nucleotide sequence information represented by SEQ ID NOS:2 to 3431;
  - (ii) classifying the isozyme identified in (i) into an isozyme having the same activity;
  - (iii) mutating all genes encoding the isozyme having the same activity simultaneously; and
  - (iv) examining productivity by a fermentation method of the compound selected in (i) of the coryneform bacterium which have been transformed with the gene obtained in (iii).
  - 59. A method for breeding a coryneform bacterium using the nucleotide sequence information represented by SEQ ID NOS:2 to 3431, comprising the following:
    - (i) arranging a function information of an open reading frame (ORF) represented by SEQ ID NOS:2 to 3431;
    - (ii) allowing the arranged ORF to correspond to an enzyme on a known biosynthesis or signal transmission pathway;
    - (iii) explicating an unknown biosynthesis pathway or signal transmission pathway of a coryneform bacterium in combination with information relating known biosynthesis pathway or signal transmission pathway of a coryneform bacterium:
    - (iv) comparing the pathway explicated in (iii) with a biosynthesis pathway of a target useful product; and
    - (v) transgenetically varying a coryneform bacterium based on the nucleotide sequence information to either strengthen a pathway which is judged to be important in the biosynthesis of the target useful product in (iv) or weaken a pathway which is judged not to be important in the biosynthesis of the target useful product in (iv).
- 35 **60.** A coryneform bacterium, bred by the method of any one of claims 52 to 59.
  - **61.** The coryneform bacterium according to claim 60, which is a microorganism belonging to the genus *Corynebacterium*, the genus *Brevibacterium*, or the genus *Microbacterium*.
- 40 62. The coryneform bacterium according to claim 61, wherein the microorganism belonging to the genus Corynebacterium is selected from the group consisting of Corynebacterium glutamicum, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum, Corynebacterium callunae, Corynebacterium herculis, corynebacterium lilium, Corynebacterium melassecola, Corynebacterium thermoamino genes, and Corynebacterium ammonia genes.
  - 63. A method for producing at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid and an analogue thereof, comprising:
    - culturing a coryneform bacterium of any one of claims 60 to 62 in a medium to produce and accumulate at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogues thereof; recovering the compound from the culture.
  - 64. The method according to claim 63, wherein the compound is L-lysine.
  - 65. A method for identifying a protein relating to useful mutation based on proteome analysis, comprising the following:
    - (i) preparing

5

15

20

25

30

45

50

a protein derived from a bacterium of a production strain of a coryneform bacterium which has been subjected to mutation breeding by a fermentation process so as to produce at least one compound selected from an amino acid, a nucleic acid, a vitamin, a saccharide, an organic acid, and analogues thereof, and a protein derived from a bacterium of a parent strain of the production strain;

- (ii) separating the proteins prepared in (i) by two dimensional electrophoresis;
- (iii) detecting the separated proteins, and comparing an expression amount of the protein derived from the production strain with that derived from the parent strain;
- (iv) treating the protein showing different expression amounts as a result of the comparison with a peptidase to extract peptide fragments;
- (v) analyzing amino acid sequences of the peptide fragments obtained in (iv); and
- (vi) comparing the amino acid sequences obtained in (v) with the amino acid sequence represented by SEQ
- ID NOS:3502 to 7001 to identifying the protein having the amino acid sequences.
- 66. The method according to claim 65, wherein the coryneform bacterium is a microorganism belonging to the genus corynebacterium, the genus Brevibacterium, or the genus Microbacterium.
  - 67. The method according to claim 66, wherein the microorganism belonging to the genus Corynebacterium is selected from the group consisting of Corynebacterium glutamicum, Corynebacterium acetoacidophilum, Corynebacterium acetoglutamicum, Corynebacterium callunae, Corynebacterium herculis, Corynebacterium lilium, Corynebacterium melassecola, Corynebacterium thermoaminogenes, and Corynebacterium ammoniagenes.
  - 68. A biologically pure culture of Corynebacterium glutamicum AHP-3 (FERM BP-7382).

10

15

20

25

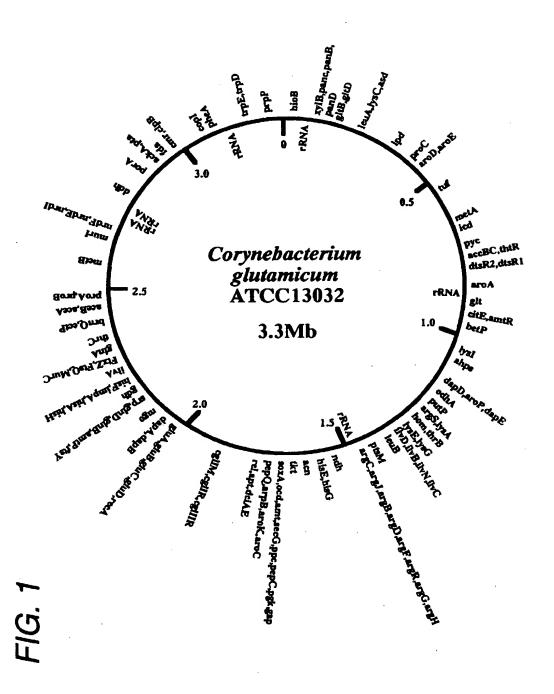
30

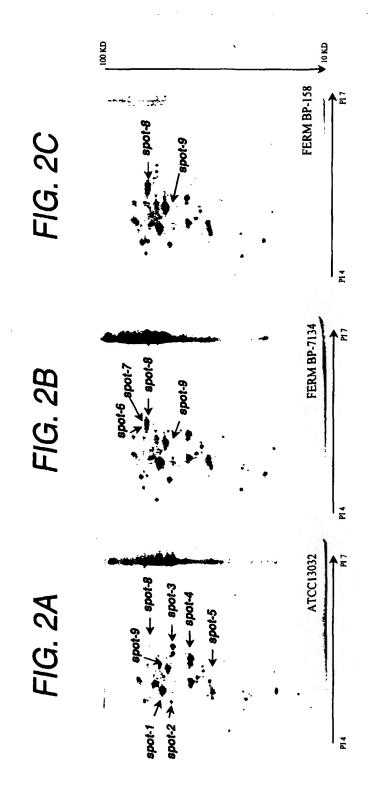
35

40

45

50





GENOME AND/OR SEQUENCE DATA POLYNUCLEOTIDE
AND/OR
POLYPEPTIDE
SEQUENCE
RECOGNIZER AND
QUERY BUFFER NETWORK REQUEST RESULT USER INPUT DEVICE

245

FIG. 4

